



**Ricardo Francisco Santos Afonso**

## **Avaliação por Detecção Remota do Efeito das Operações de Limpeza nas Faixas de Gestão de Combustível de Incêndios**

Dissertação para obtenção do Grau de Mestre em  
**Engenharia Informática**

Orientador: Carlos Augusto Isaac Piló Viegas Damásio,  
Professor, Universidade Nova de Lisboa  
Co-orientador: João Carlos Gomes Moura Pires, Professor,  
Universidade Nova de Lisboa

Júri

Presidente: Pedro Abílio Duarte de Medeiros  
Arguente: Ana Cristina Navarro Ferreira  
Vogal: Carlos Augusto Isaac Piló Viegas Damásio



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE NOVA DE LISBOA

**Setembro, 2019**



## **Avaliação por Detecção Remota do Efeito das Operações de Limpeza nas Faixas de Gestão de Combustível de Incêndios**

Copyright © Ricardo Francisco Santos Afonso, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.



## AGRADECIMENTOS

Em primeiro lugar, gostava de agradecer ao meu orientador o Professor Carlos Viegas Damásio e ao meu co-orientador o Professor João Moura Pires por todo o apoio prestado e por me terem guiado neste projeto. Foi um prazer trabalhar convosco.

Em segundo lugar, um agradecimento à Câmara Municipal de Mação por ter disponibilizado informação que foi fundamental para este trabalho.

Queria também agradecer aos meus colegas que me acompanharam neste percurso, em particular ao André Neves e Igor Ruivo, com quem debati ideias e aprendi ao longo deste ano.

Gostava também de agradecer aos meus amigos que me acompanharam ao longo deste curso e que enriqueceram a minha experiência na faculdade.

Por último, e não menos importante, queria agradecer à minha família pelo apoio que me deram ao longo destes anos, e que sem eles nada disto teria sido possível.



## RESUMO

---

Os incêndios todos os anos assolam o território Português e são uma preocupação para as entidades públicas e para a população. Para impedir a progressão dos fogos e minimizar o seu impacto, no Decreto-Lei n.º 124/2006, foram estipuladas faixas de gestão de combustível de incêndios (FGCI) ao redor de habitações ou aglomerados habitacionais, ao longo de estradas nacionais e outras infraestruturas.

A gestão das FGCI requer que haja monitorização do estado da vegetação de forma a proceder atempadamente à manutenção e limpeza destas faixas. Uma vez que o estado da vegetação é altamente influenciado por fatores climáticos e está em constante mudança, a monitorização das FGCI tem de ser feita periodicamente e requer a colaboração entre várias entidades.

De forma melhorar a monitorização das FGCI é proposto o uso de imagens de satélites, como os Sentinel-1 e Sentinel-2 em junção com índices de vegetação a fim de extrair características temporais associadas à vegetação dentro das FGCI de forma a determinar se estas foram intervencionadas.

Através destas características e com recurso a algoritmos de aprendizagem automática foram criados modelos que permitem estimar se uma determinada FGCI foi ou não intervencionada.

**Palavras-chave:** Deteção Remota, Sentinel-2, Sentinel-1, Índices de Vegetação, Séries Temporais, Aprendizagem Automática ...

---





## ABSTRACT

---

Every year the wildfires devastate the Portuguese territory and are a concern for the public entities and for the population. In order to prevent wildfire progression and minimize its impact, fire fuel management bands (FGCI) have been stipulated, in Decree-Law n °124/2006, around buildings or housing clusters, along national roads and other infrastructures.

FGCI management requires monitoring the vegetation condition in order to promptly proceed with the maintenance and cleaning of these bands. Since the vegetation status is highly influenced by climacteric factors and is constantly changing, the monitoring of FGCI has to be done periodically and requires the collaboration of multiple entities.

In order to improve FGCI monitoring, we propose the use of satellite images, such as the Sentinel-1 and Sentinel-2, along with vegetation indices in order to extract temporal characteristics associated with the vegetation within the FGCI to determine if they were intervened.

Through these characteristics and using machine learning algorithms, models were developed to estimate if an FGCI was intervened.

**Keywords:** Remote Detection, Sentinel-2, Sentinel-1, Vegetation Indices, Time Series, Machine Learning...

---



# ÍNDICE

<b>Lista de Figuras</b>	<b>xv</b>
<b>Lista de Tabelas</b>	<b>xxi</b>
<b>Siglas</b>	<b>xxiii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Descrição do Problema . . . . .	2
1.3 Área de Estudo . . . . .	4
1.4 Objetivo e Proposta de Solução . . . . .	5
1.5 Contribuições . . . . .	6
1.6 Estrutura do Documento . . . . .	6
<b>2 Conceitos e Trabalhos Relacionados</b>	<b>7</b>
2.1 Introdução . . . . .	7
2.2 Faixas de Gestão de Combustíveis de Incêndios . . . . .	8
2.2.1 Redes de Faixas de Gestão de Combustíveis . . . . .	8
2.2.2 Rede Primária de Faixas de Gestão de combustíveis (RPFGC) . . . . .	9
2.2.3 Rede Secundária e Rede Terciária . . . . .	9
2.2.4 Informação de Referência das FGCI . . . . .	11
2.3 Detecção Remota . . . . .	13
2.3.1 Satélites Multiespectrais . . . . .	14
2.3.2 Synthetic-Aperture Radar (SAR) . . . . .	15
2.3.3 Pré-processamento de Imagens Satélite . . . . .	15
2.4 Índices de Vegetação . . . . .	17
2.4.1 <i>Normalized Difference Vegetation Index</i> . . . . .	17
2.4.2 <i>Soil-Adjusted Vegetation Index</i> . . . . .	18
2.4.3 <i>Simple Ratio</i> . . . . .	19
2.4.4 <i>Inverted Red-Edge Chlorophyll Index</i> . . . . .	20
2.4.5 <i>Normalized Difference Water Index</i> . . . . .	20
2.5 Trabalhos Relacionados . . . . .	21
2.5.1 Classificação de Vegetação . . . . .	22

2.5.2	Monitorização de Culturas . . . . .	23
2.5.3	Outros Trabalhos . . . . .	23
2.5.4	Pontos Essenciais . . . . .	24
2.6	Conclusões . . . . .	24
<b>3</b>	<b>Classificação Automática em Detecção Remota</b>	<b>27</b>
3.1	Introdução . . . . .	27
3.2	Algoritmos de classificação . . . . .	27
3.2.1	<i>K-Nearest Neighbors</i> . . . . .	28
3.2.2	<i>Support Vector Machine</i> . . . . .	28
3.2.3	<i>Random Forest</i> . . . . .	29
3.2.4	<i>Extreme Gradient Boosting</i> . . . . .	30
3.3	Abordagens de classificação . . . . .	31
3.4	Seleção de Características . . . . .	31
3.5	Validação . . . . .	32
3.5.1	Matriz de Confusão . . . . .	33
3.5.2	Precisão . . . . .	33
3.5.3	Recall . . . . .	34
3.5.4	F1-score . . . . .	34
3.5.5	Accuracy . . . . .	34
3.5.6	Kappa . . . . .	34
3.6	Conclusões . . . . .	35
<b>4</b>	<b>Abordagem</b>	<b>37</b>
4.1	Introdução . . . . .	37
4.2	Pré-processamento de Imagens Satélite . . . . .	38
4.2.1	Sentinel-1 . . . . .	38
4.2.2	Sentinel-2 . . . . .	39
4.2.3	Reamostragem de Imagens . . . . .	39
4.3	Determinação das Zonas de Interesse das FGCI . . . . .	40
4.3.1	FGCI . . . . .	40
4.3.2	<i>Buffers</i> . . . . .	40
4.3.3	Divisão em Secções . . . . .	42
4.4	Conjunto de Dados . . . . .	43
4.5	Análise Estática . . . . .	45
4.6	Análise Temporal . . . . .	45
4.7	Conclusões . . . . .	46
<b>5</b>	<b>Implementação</b>	<b>47</b>
5.1	Introdução . . . . .	47
5.2	Alinhamento de Imagens . . . . .	47
5.3	Informação vetorial . . . . .	49

5.3.1	Divisão das FGCI por Tipo e União de Polígonos . . . . .	49
5.3.2	Criação de Buffers . . . . .	50
5.3.3	Remoção de Áreas Sobrepostas e de Fragmentos . . . . .	55
5.3.4	Divisão das FGCI em secções . . . . .	56
5.4	Extração de Dados . . . . .	57
5.5	Conclusões . . . . .	58
<b>6</b>	<b>Trabalho Experimental</b>	<b>59</b>
6.1	Introdução . . . . .	59
6.2	Especificações de Hardware . . . . .	59
6.3	Dados para Classificação . . . . .	60
6.4	Afinação de Parâmetros . . . . .	61
6.4.1	KNN . . . . .	61
6.4.2	SVM . . . . .	62
6.4.3	RF . . . . .	62
6.4.4	XGBoost . . . . .	63
6.5	Análise Estática . . . . .	63
6.5.1	Agrupamento . . . . .	64
6.5.2	Classificação Supervisionada . . . . .	67
6.6	Análise Temporal . . . . .	70
6.6.1	Classificação Através de Imposição de Limite . . . . .	72
6.6.2	Classificação Supervisionada . . . . .	73
6.7	Importância das Características . . . . .	81
6.7.1	Faixas ao Longo de Estradas . . . . .	81
6.7.2	Faixas ao Longo de Linhas Elétricas . . . . .	83
6.8	Estimativa de Datas de Intervenção . . . . .	85
6.9	Discussão dos Resultados . . . . .	87
6.9.1	Classificação das Intervenções nas FGCI . . . . .	87
6.9.2	Seleção de características . . . . .	90
6.9.3	Datas de Intervenção . . . . .	90
<b>7</b>	<b>Conclusão e Trabalho Futuro</b>	<b>93</b>
7.1	Conclusão . . . . .	93
7.2	Trabalho Futuro . . . . .	94
	<b>Bibliografia</b>	<b>97</b>
<b>I</b>	<b>Anexo A</b>	<b>103</b>



## LISTA DE FIGURAS

1.1	Informação vetorial de algumas faixas na freguesia de Cardigos, concelho de Mação. . . . .	3
1.2	Zona de estudo e algumas faixas em formato vetorial na freguesia de Cardigos, Mação. . . . .	4
2.1	Exemplo de uma FRC na Nazaré. . . . .	9
2.2	Exemplo de uma FIC em Silves. . . . .	9
2.3	Faixa da Rede Primária na zona de Silves. . . . .	10
2.4	Linha de alta tensão. . . . .	10
2.5	FGCI ao longo de uma estrada. . . . .	10
2.6	Faixa no redor de uma habitação . . . . .	10
2.7	Exemplo de faixas mal definidas, na zona norte do distrito de Castelo Branco. (a) Imagem do satélite Sentinel-2 numa zona de habitações. (b) Polígonos representativos de faixas em redor a habitações, que foram mal definidos. . . . .	13
2.8	(a) Imagem a cores de um zona do distrito de Santarém. (b) Resultado do cálculo do NDVI para a mesma zona. . . . .	18
2.9	(a) Imagem a cores de um zona do distrito de Santarém (b) Resultado do cálculo do SAVI para a mesma zona. . . . .	18
2.10	(a) Imagem a cores de um zona do distrito de Santarém (b) Resultado do cálculo do SR para a mesma zona. . . . .	19
2.11	(a) Imagem a cores de um zona do distrito de Santarém (b) Resultado do cálculo do IRECI para a mesma zona. . . . .	20
2.12	(a) Imagem a cores de um zona do distrito de Santarém (b) Resultado do cálculo do NDWI para a mesma zona. . . . .	21
3.1	Exemplo do KNN com valores de K=1 e K=3. (fonte: <a href="https://medium.com/datadriveninvestor/knn-algorithm-and-implementation-from-scratch-b9f9b739c28f">https://medium.com/datadriveninvestor/knn-algorithm-and-implementation-from-scratch-b9f9b739c28f</a> ) . . . . .	28
3.2	Exemplo de uma separação linear do SVM. (fonte: <a href="https://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html?highlight=svm">https://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html?highlight=svm</a> ) . . . . .	29

3.3	Exemplo do RF e das suas árvores de decisão. (fonte: <a href="https://medium.com/@ar.ingenious/applying-random-forest-classification-machine-learning-algorithm-from-scratch-with-real-24ff198a1c57">https://medium.com/@ar.ingenious/applying-random-forest-classification-machine-learning-algorithm-from-scratch-with-real-24ff198a1c57</a> ) . . . . .	30
3.4	Faixas ao longo de estradas e de linhas elétricas, no norte de Mação, que foram intervencionadas no ano 2018. . . . .	32
4.1	Diagrama que representa os passos da abordagem seguida. . . . .	37
4.2	Reamostragem da banda B07 do Sentinel-2 de 20m para 10m. (a) Imagem com 20m de resolução. (b) Imagem reamostrada, com 10m de resolução. . . . .	39
4.3	Exemplos de <i>buffers</i> para os diferentes tipos de faixas. . . . .	41
4.4	Divisão das FGCI ao longo das estradas e dos seus <i>buffers</i> em secções . . . . .	42
4.5	Exemplos da intersecção das FGCI e dos <i>buffers</i> com as secções criadas. (a) Secção de uma FGCI ao longo de uma estrada. (b) Secção de uma FGCI ao longo de uma linha elétrica. (c) Secção de uma FGCI ao redor de um aglomerado de habitações. . . . .	43
5.1	Correspondência de pontos entre duas imagens semelhantes, sendo que a imagem da esquerda apenas tem uma rotação. . . . .	48
5.2	Cadeia de passos para o processamento das FGCI e criação dos <i>buffers</i> . . . . .	49
5.3	Exemplo do resultado da união de polígonos (faixas ao longo de estradas). (a) Formato original da FGCI. (b) FGCI após união de polígonos. . . . .	50
5.4	Fases da criação dos <i>buffers</i> exteriores. (a) Polígono que representa uma FGCI ao longo de linhas elétricas. (b) Polígono resultante do uso da função <i>Buffer</i> com 20m. (c) Polígonos que representam apenas áreas exteriores à FGCI. . . . .	52
5.5	Exemplo do resultado da aplicação da função <i>Buffer</i> em algumas faixas ao redor de habitações. (a) Polígonos originais que representam a FGCI. (b) Polígonos resultantes após a criação de um <i>buffer</i> de 20m. . . . .	53
5.6	Processo para obter o interior das FGCI ao redor de habitações. . . . .	54
5.7	Resutaldo da diferença entre a FGCI com um buffer de 20m (Figura 5.5b) e o polígono que representa o interior da FGCI (Figura 5.6d). . . . .	55
5.8	Exemplo da remoção de áreas de sobreposição com outras faixas. (a) Sobreposição do <i>buffer</i> com FGCI ao longo de estradas. (b) <i>buffer</i> resultante da remoção de zonas de sobreposição. . . . .	55
5.9	Fases da criação de secções. (a) Pontos do interior (laranja) e do <i>buffer</i> exterior (cinzento) da FGCI. (b) <i>Clusters</i> gerados usando K-Means. (c) Polígonos gerados para cada um dos <i>clusters</i> . . . . .	56
5.10	Tempos de recorte das imagens por FGCI e por satélite. . . . .	57
6.1	Datas das imagens usadas de ambos os satélites. . . . .	60
6.2	(a) NDVI calculado usando imagens do Sentinel-2 de 25-04-2018. (b) NDVI calculado usando imagens do Sentinel-2 de 15-05-2018. . . . .	64



6.3	(a) Secções consideradas para realizar a classificação. (b) Secções que foram intervencionadas (branco) e não intervencionadas (vermelho). . . . .	65
6.4	(a) Faixas faixas ao redor de habitações na zona analisada. (b) Resultado do agrupamento das secções de faixas ao redor de habitações. . . . .	66
6.5	F1-score para os diferentes algoritmos e conjuntos de dados nas faixas ao longo das estradas. . . . .	67
6.6	Kappa para os diferentes algoritmos e conjuntos de dados nas faixas ao longo das estradas. . . . .	68
6.7	F1-score para os diferentes algoritmos e conjuntos de dados, nas faixas ao longo das linhas elétricas. . . . .	69
6.8	Kappa para os diferentes algoritmos e conjuntos de dados, nas faixas ao longo das linhas elétricas. . . . .	69
6.9	Comportamento da $diff_{NDVI}$ ao longo de 2018 em secções que foram intervencionadas entre 25-04-2018 e 15-05-2018 (a) e em secções que não foram intervencionadas (b). . . . .	71
6.10	Valor das métricas F1-score, <i>recall</i> e precisão para os diferentes valores do limite, na classificação de secções intervencionadas (a) e não intervencionadas (b) usando o índice NDVI. . . . .	72
6.11	Valor da métrica F1-score para os diferentes algoritmos e conjuntos de dados com dados do Sentinel-2. . . . .	74
6.12	Exemplos de falsos dos positivos registados com o SVM e conjunto de dados <i>CD_TUDO</i> . . . . .	74
6.13	Valor das métricas F1-score e Kappa para os diferentes algoritmos para os dados do Sentinel-1. . . . .	75
6.14	Valor da métrica F1-score para os diferentes algoritmos e conjuntos de dados, com dados do Sentinel-1 e Sentinel-2. . . . .	76
6.15	Valor da métrica F1-score para os diferentes algoritmos e conjuntos de dados. . . . .	77
6.16	Exemplos de falsos dos positivos registados com o XGBoost e conjunto de dados <i>CD_BANDAS</i> . . . . .	78
6.17	Valores das métricas F1-score e Kappa para os diferentes algoritmos usando os dados do Sentinel-1. . . . .	79
6.18	Valor da métrica F1-score para os diferentes algoritmos e conjuntos de dados com dados do Sentinel-1 e Sentinel-2. . . . .	80
6.19	Número de características selecionadas e eliminadas pelo LASSO, agrupadas pelo tipo de dados (Bandas e índices de vegetação) obtidos do Sentinel-2. . . . .	81
6.20	Número de características selecionadas e eliminadas pelo LASSO, agrupadas pelo tipo de característica. . . . .	82
6.21	F1-score antes e depois de ter sido realizada a seleção de características. . . . .	82
6.22	Tempos de afinação dos parâmetros dos algoritmos, antes e depois da seleção de características. . . . .	83

6.23	Número de características selecionadas e eliminadas pelo LASSO, agrupadas pelo tipo de dados (Bandas e índices de vegetação) obtidos do Sentinel-2. . . .	83
6.24	Número de características selecionadas e eliminadas pelo LASSO, agrupadas pelo tipo de característica. . . . .	83
6.25	F1-score antes e depois de ter sido realizada a seleção de características. . . .	84
6.26	Tempos de afinação dos parâmetros dos algoritmos, antes e depois da seleção de características. . . . .	84
6.27	Secções usadas para estimar a data de intervenção e que foram intervencionadas na mesma altura do ano. . . . .	85
6.28	NDVI antes e depois da intervenção para o Local A. (a) NDVI dia 15-05-2018. (b) NDVI dia 19-06-2018. . . . .	85
6.29	NDVI antes e depois da intervenção para o Local B. (a) NDVI dia 15-05-2018. (b) NDVI dia 19-06-2018. . . . .	86
6.30	Resultados da previsão das datas para o Local A (a) e para o Local B (b). A vermelho secções onde a estimativa foi errada, a verde onde foi correta, e a preto secções que não obtiveram resultados. . . . .	86
I.1	Métricas da análise estática para faixas ao longo de estradas. . . . .	104
I.2	Matriz de confusão da análise estática para faixas ao longo de estradas. . . .	105
I.3	Métricas da análise estática para faixas ao longo de linhas elétricas. . . . .	106
I.4	Matriz de confusão da análise estática para faixas ao longo de linhas elétricas. . . . .	107
I.5	Métricas da análise temporal para faixas ao longo de estradas usando dados do Sentinel-2. . . . .	108
I.6	Matriz de confusão da análise temporal para faixas ao longo de estradas usando dados do Sentinel-2. . . . .	109
I.7	Métricas da análise temporal para faixas ao longo de estradas usando dados do Sentinel-1. . . . .	109
I.8	Matriz de confusão da análise temporal para faixas ao longo de estradas usando dados do Sentinel-1. . . . .	110
I.9	Métricas da análise temporal para faixas ao longo de estradas usando dados do Sentinel-2 e Sentinel-1. . . . .	111
I.10	Matriz de confusão da análise temporal para faixas ao longo de estradas usando dados do Sentinel-2 e Sentinel-1. . . . .	112
I.11	Métricas da análise temporal para faixas ao longo de linhas usando dados do Sentinel-2. . . . .	113
I.12	Matriz de confusão da análise temporal para faixas ao longo de linhas usando dados do Sentinel-2. . . . .	114
I.13	Métricas da análise temporal para faixas ao longo de linhas usando dados do Sentinel-1. . . . .	114
I.14	Matriz de confusão da análise temporal para faixas ao longo de linhas usando dados do Sentinel-1. . . . .	115

I.15 Métricas da análise temporal para faixas ao longo de linhas usando dados do Sentinel-2 e Sentinel-1. . . . .	116
I.16 Matriz de confusão da análise temporal para faixas ao longo de linhas usando dados do Sentinel-2 e e Sentinel-1. . . . .	117



## LISTA DE TABELAS

2.1	Atributos associados a cada uma das faixas e a respetiva descrição [32]. . . .	12
2.2	Bandas disponibilizadas pelo Sentinel-2A (S2A) e Sentinel-2B (S2B) [42]. . .	14
2.3	Descrição dos produtos do Sentinel-2 que são disponibilizados [41]. . . . .	15
3.1	Exemplo da estrutura de uma matriz de confusão para duas classes. . . . .	33
4.1	Características calculadas a partir dos dados de cada um dos satélites. . . . .	43
4.2	Conjuntos de dados usados e respetivas características. . . . .	44
6.1	Número de secções por tipo de FGCI e a média de área das secções. . . . .	60
6.2	Número de características de cada conjunto de dados para os tipos de análise realizados. . . . .	61
6.3	Valores usados para os parâmetros do KNN. . . . .	62
6.4	Valores usados para os parâmetros do SVM. . . . .	62
6.5	Valores usados para os parâmetros do RF. . . . .	63
6.6	Valores usados para os parâmetros do XGBoost. . . . .	63
6.7	Matriz de confusão dos resultados do agrupamento realizado. . . . .	66
6.8	Matriz de confusão para o algoritmo XGBoost usando o conjunto de dados <i>CD_TUDO</i> . . . . .	68
6.9	Matriz de confusão para o algoritmo RF usando o conjunto de dados <i>CD_TUDO</i> . . . . .	70
6.10	Matriz de confusão para o algoritmo SVM usando o conjunto de dados <i>CD_TUDO</i> . . . . .	74
6.11	Matriz de confusão para o algoritmo RF para os dados do Sentinel-1. . . . .	75
6.12	Matriz de confusão para o algoritmo XGBoost usando o conjunto de dados <i>CD_BANDAS</i> . . . . .	77
6.13	Matriz de confusão para o algoritmo RF usando os dados das polarizações do Sentinel-1. . . . .	79
6.14	Matriz de confusão para o algoritmo XGBoost usando o conjunto de dados <i>CD_TUDO</i> . . . . .	80



## SIGLAS

ESA	<i>European Sapce Agency.</i>
FGCI	Faixas de Gestão de Combustível de Incêndios.
FIC	Faixas de Interrupção de Combustível.
FN	Falso Negativo.
FP	Falso Positivo.
FRC	Faixas de Redução de Combustível.
GRD	<i>Ground Range Detected.</i>
GRI	<i>Imagem de referência Global.</i>
ICNF	<i>Instituto da Conservação da Natureza e das Florestas .</i>
IRECI	<i>Inverted Red-Rdge Chlorophyll Index.</i>
IW	<i>Interferometric Wide.</i>
KNN	<i>K-Nearest Neighbors.</i>
LAI	<i>Leaf Area Index.</i>
NDVI	<i>Normalized Difference Vegetation Index.</i>
NDWI	<i>Normalized Difference Water Index .</i>
NIR	<i>Near-Infrared.</i>
PCA	<i>Principal Component Analysis .</i>
PMDFCI	Plano Municipal de Defesa da Floresta Contra os Incêndios.
RBF	<i>Radial Basis Function.</i>
RE	<i>Red-Edge.</i>
RF	<i>Random Forest.</i>

RPFGC     Rede Primária de Gestão de Combustível.

SAR        *Synthetic-Aperture Radar .*

SAVI       *Soil-Adjusted Vegetation Index .*

SLC        *Single Look Complex.*

SNAP       *Sentinel Application Platform.*

SR          *Simple Ratio .*

SVM        *Support Vector Machine .*

SWIR       *Short Wave Infrared.*

TWDTW    *Time-Weighted Dynamic Time Warping .*

VN          Verdadeiro Negativo.

VP          Verdadeiro Positivo.

XGBoost    *Extreme Gradient Boosting.*





## INTRODUÇÃO

Os incêndios florestais constituem uma ameaça para o ecossistema, bens e populações. Todos os anos em Portugal, milhares de hectares de floresta são consumidos pelas chamas [4] causando danos ambientais, materiais e pessoais.

Com o propósito de minimizar os danos causados pelos incêndios foi criada legislação e estabelecido o Sistema Nacional de Defesa da Floresta Contra Incêndios (SNDFCI), que estipula a criação de Faixas de Gestão de Combustível de incêndios (FGCI) a nível nacional, distrital e municipal [3, 23].

As FGCI são faixas florestais nas quais as árvores e a vegetação curta (ervas e arbustos) foram reduzidas ou completamente removidas, criando uma área menos propícia ao alastramento de fogo. São faixas estrategicamente planeadas para proteger bens e infraestruturas, facilitar o combate e limitar a propagação do fogo. Para tal estas faixas têm de respeitar um conjunto de condições que determinam fatores como a dimensão, densidade da vegetação, áreas e património a proteger.

Para que as FGCI cumpram o seu propósito é necessário que haja monitorização da vegetação dentro do perímetro das faixas, permitindo, caso necessário, que seja efetuada atempadamente uma limpeza.

### 1.1 Motivação

A gestão de vegetação nas FGCI é uma tarefa que requer a colaboração de entidades públicas e privadas, garantindo sempre que as FGCI obedecem às condições exigidas na lei [3, 23].

O estado da vegetação nestas faixas está dependente de fatores climáticos que podem, num período de dias, provocar mudanças significativas na vegetação. Como consequência é necessário que o processo de monitorização das FGCI seja realizado continuamente de modo a que haja uma intervenção no momento adequado.

Uma vez que as FGCI são de grande extensão, a existência de processos automáticos ou semi-automáticos que permitam identificar se uma determinada faixa foi intervencionada, faria com que fosse possível obter informação atualizada sobre as condições da vegetação nestas faixas. Estes processos facilitariam a monitorização, acelerando a deteção de faixas com necessidade de intervenção, permitindo assim uma manutenção atempada.

Além disso, este tipo de informação pode também ser útil no planeamento ao combate de incêndios, indicando áreas onde o combate poderá ser facilitado devido às condições da vegetação das FGCI.

A existência de dados de observações satélite com resoluções que vão até aos 10 x 10 metros, disponíveis publicamente, abre a possibilidade da integração de novos métodos, que recorram à deteção remota, no processo de monitorização das FGCI.

Contudo, apesar de existirem variados trabalhos que abordam a temática de monitorização ou classificação de vegetação e culturas, até à data não foram identificados trabalhos que tenham como foco a vegetação nas FGCI.

Assim, nesta tese são usadas técnicas de deteção remota aplicadas ao contexto específico das FGCI de forma a analisar a vegetação dentro destas faixas e determinar se foram intervencionadas.

### 1.2 Descrição do Problema

As FGCI todos os anos têm de ser intervencionadas para que a quando da época de incêndios a vegetação no seu interior seja o mais reduzida possível, de forma a que estas faixas cumpram o seu propósito.

A elevada extensão destas faixas e o facto de grande parte das FGCI intersectar com terrenos de particulares pode dificultar o processo de limpeza e gerar descontinuidades nas áreas intervencionadas das FGCI. Estes fatores fazem com que a monitorização das faixas intervencionadas ou de deteção de faixas com necessidade de ser intervencionadas seja um problema complexo.

É importante haver uma distinção entre intervenção e limpeza pois uma faixa pode se encontrar limpa por diversas razões (e.g. incêndios, campos agrícolas, etc) e não ter sido intervencionada. Assim, para detetar uma intervenção propõe-se a comparação da vegetação presente no interior das FGCI e com a vegetação nas zonas exteriores adjacentes (*buffers*) à faixa.

Para concretizar esta abordagem de comparação de vegetação é necessário criar *buffers* adjacentes aos limites das FGCI. O processo de criação destes *buffers* não é trivial pois as FGCI têm diferentes dimensões (podem variar entre os 10m e os 125m de largura), têm um formato que varia de acordo com o tipo de faixa e além disso é necessário garantir que

estes *buffers* se encontram sobre zonas de vegetação, sendo preciso remover intersecções com estradas ou com outras FGCI.

Na Figura 1.1 podem observar-se alguns exemplos de FGCI em formato vetorial. Neste caso em particular as faixas ao longo das estradas têm 15m de largura, assim como as faixas ao longo de linhas elétricas. Já as faixas que pertencem à Rede Primária de Faixas de Gestão de Combustível (RPFGC) (rede de faixas a nível distrital que cobre parte do território português) têm uma largura de 125m. No caso das faixas ao redor de habitações estas têm 50m em habitações isoladas e 100m em aglomerados habitacionais.

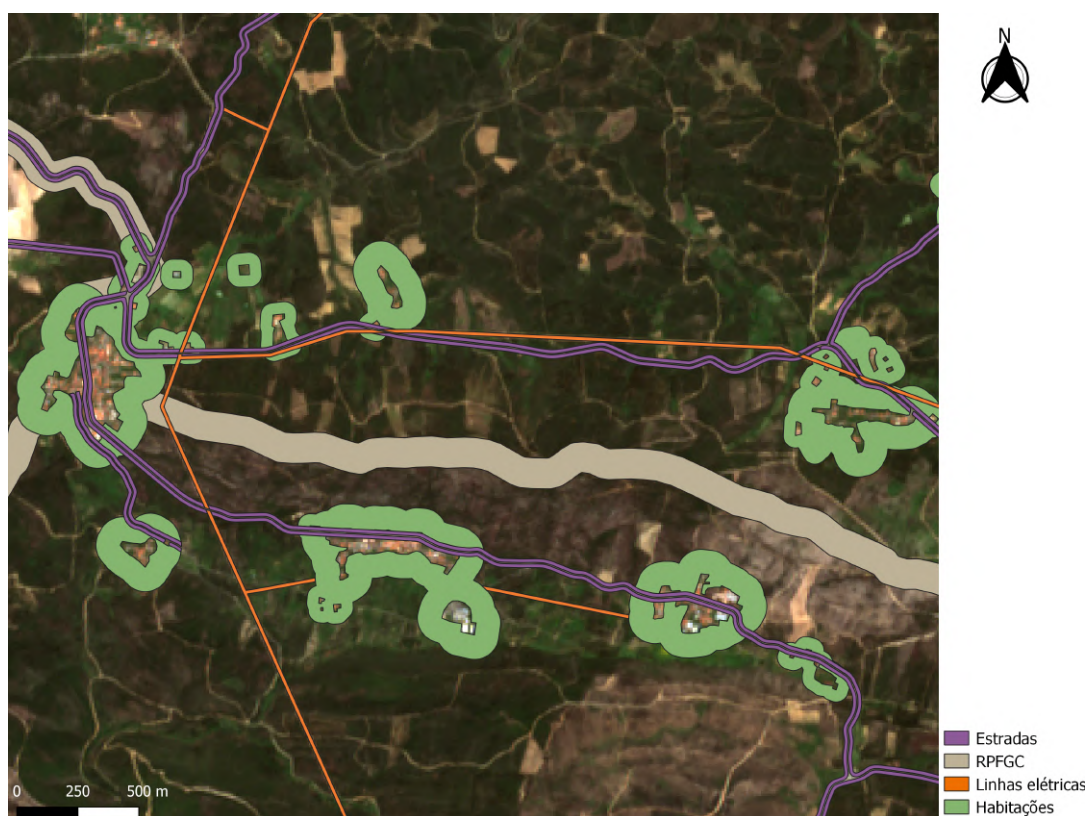


Figura 1.1: Informação vetorial de algumas faixas na freguesia de Cardigos, concelho de Mação.

A disposição da vegetação dentro destas faixas é diferente consoante o tipo de faixa. Para a maioria das faixas é requerido por lei a remoção total de todos os tipos de combustível, mas existem alguns casos em que apenas é requerido a redução do material combustível, permitindo a existência de alguma vegetação como árvores ou arbustos.

Por outro lado, existem faixas que a sua principal função não é impedir a propagação do fogo ou proteger infraestruturas, mas sim impedir a sua ignição, como é o caso das faixas ao longo de linhas elétricas.

Neste trabalho são usados satélites do programa Copernicus, que recolhem observações e disponibilizam publicamente. O Sentinel-2A e o Sentinel-2B (lançados em 2015 e 2017, respetivamente), são dois satélites deste programa, que fornecem bandas no visível e no infravermelho próximo com resoluções que vão até aos 10 x 10 metros.

Estes satélites captam imagens com resoluções espaciais suficientemente finas para que seja possível detetar mudanças na vegetação nas áreas florestais das FGCI. Porém, em alguns casos a resolução espacial destes satélite é muito próxima (e em alguns casos igual) à largura de algumas faixas, como é o caso das faixas ao longo das estradas. Isto pode apresentar algumas dificuldades, nomeadamente quando a precisão geográfica das imagens não é a melhor, fazendo com que o vetor que representa a FGCI não corresponda com a localização da FGCI na imagem.

Todos estes aspetos que estão associados às FGCI fazem com que a deteção de intervenções na vegetação destas faixas seja uma tarefa complexa e que necessita de uma abordagem cuidadosa.

### 1.3 Área de Estudo

Como área de estudo foi escolhido o concelho de Mação, que faz parte do distrito de Santarém. Este concelho já foi fustigado várias vezes pelos incêndios e realiza há vários anos a gestão da vegetação dentro das FGCI.

Foi usada informação em formato vetorial das faixas deste concelho. O Instituto de Conservação da Natureza e Florestas (ICNF) disponibiliza esta informação para todos os concelhos de Portugal continental. Na Figura 1.2 pode ser visto um exemplo de algumas faixas em formato vetorial na zona norte do concelho de Mação.

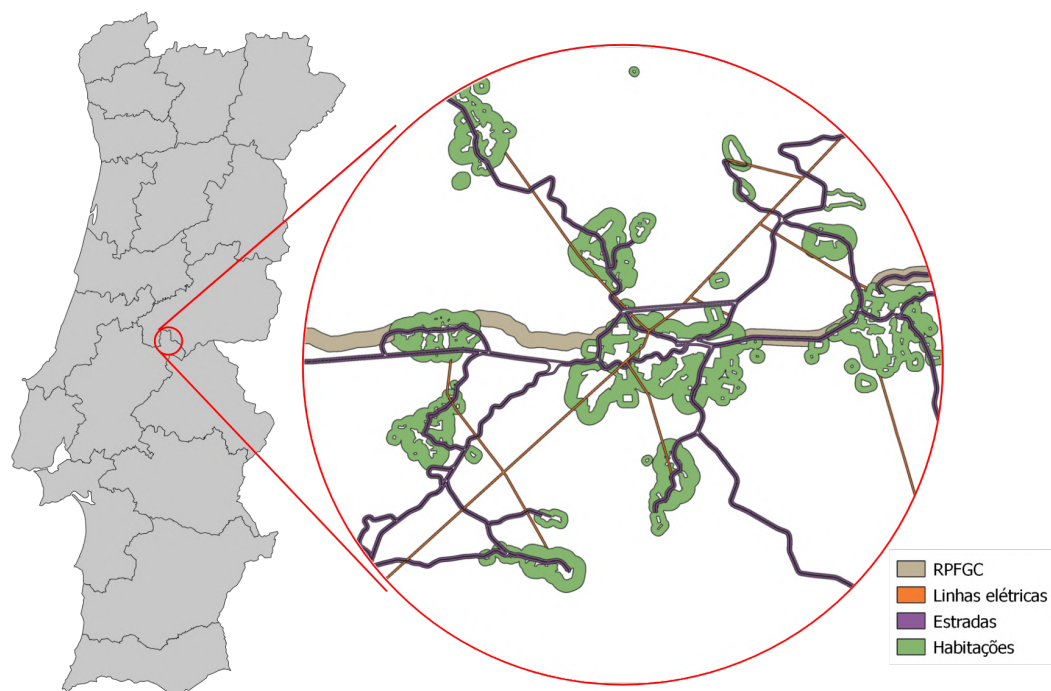


Figura 1.2: Zona de estudo e algumas faixas em formato vetorial na freguesia de Cardigos, Mação.

Além da informação vetorial foram estabelecidos acordos entre o NOVA LINCS (laboratório de investigação nas áreas da Ciência da Computação e da Informática, situado na Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa) e a câmara municipal de Mação, que permitiram obter informação atualizada sobre as faixas que foram intervencionadas no ano de 2018.

Neste concelho, e no ano em questão, foram intervencionadas faixas ao longo de estradas e ao longo de linhas elétricas, e como tal estas são as faixas onde a maioria das experiências foram realizadas.

### 1.4 Objetivo e Proposta de Solução

O foco deste trabalho é analisar a vegetação nas FGCI e nas zonas adjacentes a estas, recorrendo a imagens satélite, de forma a determinar se estas faixas foram intervencionadas.

Usando as imagens dos satélites Sentinel-2A e o Sentinel-2B e os valores das bandas disponibilizadas podem ser calculados índices de vegetação que possibilitam estimar características da vegetação como por exemplo: vigor da vegetação, do conteúdo de água ou densidade da vegetação.

Também através de satélites com sistema radar, como é o caso do Sentinel-1, podem ser obtidas imagens com diferentes características, mesmo em condições atmosféricas adversas, que podem também ser utilizadas para este tipo aplicação.

Usando as imagens de satélite e a informação vetorial disponibilizada pelo ICNF é possível extrair dados relativos ao interior das FGCI. Além da informação do interior das FGCI, a abordagem usada, e na qual assenta a análise feita neste trabalho, usa informação de zonas exteriores adjacentes às faixas de forma a comparar com a informação do interior das FGCI. Com estes dados e recorrendo a séries temporais podem ser identificadas características dos padrões temporais nas áreas que dizem respeito às FGCI.

Usando estas características e inserindo os seus valores em algoritmos de aprendizagem automática podemos estimar se uma faixa foi intervencionada. Contudo, a precisão destas estimativas pode variar dependendo do algoritmo usado, das características usadas ou das propriedades da faixa que está a ser analisada.

Assim, recorrendo aos diversos índices de vegetação, em combinação com as várias bandas dos satélites, e a algoritmos de aprendizagem automática é pretendida a elaboração de uma solução que permita determinar se uma determinada FGCI foi ou não intervencionada.

### 1.5 Contribuições

Do desenvolvimento desta tese resultam alguns métodos e ferramentas que foram fundamentais para concretizar a solução para o problema proposto.

É apresentada uma metodologia para proceder à deteção de intervenções nas FGCI, que recorre a informação sobre a vegetação presente no interior e no exterior próximo da faixa. É explicada a abordagem usada para realizar o pré-processamento das imagens satélite e para manipular a informação vetorial de forma a obter as áreas de interesse, sendo que esta abordagem pode ser usada para analisar qualquer tipo FGCI dos diferentes concelhos de Portugal continental.

Disto resulta um conjunto de ferramentas e código que dada informação vetorial sobre as FGCI e um conjunto de imagens satélite é capaz de extrair métricas dos padrões de séries temporais, e produzir uma estimativa das FGCI que foram ou não intervencionadas.

### 1.6 Estrutura do Documento

Este documento está dividido em vários capítulos. Nesta secção é apresentada uma breve descrição de cada um deles.

O Capítulo 1 deste documento apresenta o contexto e motivação desta dissertação, expondo numa breve introdução os temas envolventes, a descrição do problema e a proposta de solução.

No Capítulo 2 são descritos diferentes conceitos como os tipos de FGCI as suas características e respetivas funções, é feita uma descrição dos satélites usados e de alguns índices de vegetação, e são apresentados alguns trabalhos de deteção remota relacionados com a temática de monitorização e classificação de vegetação.

No Capítulo 3 deste documento, são detalhados os algoritmos de aprendizagem automática usados, e também onde são analisadas abordagens usadas na classificação de imagens satélite e respetivos métodos de validação.

No Capítulo 4 deste documento, é delineada a abordagem usada no processamento das imagens satélite, no tratamento da informação vetorial das zonas de interesse, as abordagens de classificação usadas e os respetivos conjuntos de dados.

O Capítulo 5 aborda detalhes da implementação de cada uma das fases e componentes que fazem parte da abordagem definida.

No Capítulo 6 deste documento, são apresentadas as diferentes experiências realizadas, assim como os resultados obtidos.

Finalmente, no Capítulo 7 são tiradas conclusões sobre os resultados obtidos e apresentados tópicos de trabalho futuro que assentam no trabalho realizado.





## CONCEITOS E TRABALHOS RELACIONADOS

### 2.1 Introdução

As FGCI providenciam áreas onde a densidade e distribuição da vegetação fazem com que seja difícil a propagação de fogo. Características como a dimensão e formato das faixas, assim como a distribuição da vegetação podem ter impacto na abordagem a usar para determinar se uma determinada faixa foi intervencionada.

Estas faixas apresentam várias características que podem variar consoante o tipo de faixa em questão. Como tal, são descritos aspetos de cada tipo de faixa, como as dimensões ou distribuição da vegetação, para estabelecer a base na qual será elaborada a abordagem a ser usada.

Por outro lado, uma vez que são usadas imagens satélite para obter informação sobre o interior das FGCI, é fundamental perceber as diferentes características dos satélites usados, de forma a extrair informação útil e correta para o problema em questão. Com esta informação podem ser calculados índices de vegetação, indicadores do estado da vegetação.

É também necessário analisar trabalhos relacionados com a classificação e monitorização de vegetação para perceber as técnicas e abordagens usadas na resolução de problemas relacionados com esta temática.

Tendo isto em conta, neste capítulo são abordados os conceitos base das temáticas FGCI, apresentados os satélites e alguns índices de vegetação e analisados alguns trabalhos de deteção remota relacionados com a temática de monitorização e classificação de vegetação.

## 2.2 Faixas de Gestão de Combustíveis de Incêndios

A base deste trabalho consiste na monitorização da vegetação nas FGCI. Como tal, existe uma série de critérios associados a estas faixas que têm de ser tidos em conta pois poderão ter implicações na forma como os dados serão processados e analisados.

Diferentes tipos de faixa apresentam diferentes características. É necessário perceber com algum detalhe os aspetos que envolvem a criação e a manutenção das FGCI. Características como a extensão, dimensões, período de limpeza, tipo de vegetação e a sua distribuição podem ter impacto nas séries temporais criadas a partir das bandas espectrais dos satélites.

Para estabelecer qual a melhor abordagem a seguir, quais as limitações e que cuidados a ter no processamento e na análise dos dados, temos de ter em conta estes aspetos.

Nesta secção são descritas as características que as FGCI têm de respeitar e que estão estabelecidas na legislação [3, 23].

### 2.2.1 Redes de Faixas de Gestão de Combustíveis

As FGCI cobrem grande parte do território português formando uma rede constituída por três níveis: **Rede Primária**, **Rede Secundária** e **Rede Terciária**. A rede primária é definida a nível distrital e cobre parte do território português (maioritariamente zonas do interior), enquanto que a secundária e terciária são a nível municipal e local e são definidas pelos municípios.

Cada um destes níveis é caracterizado pelas funções que cumpre. Existem três funções base que definem o propósito destes níveis:

1. Diminuição da superfície percorrida por grandes incêndios, facilitando o combate ao fogo;
2. Redução dos efeitos de incêndios, protegendo vias de comunicação, infraestruturas e equipamentos, zonas edificadas e povoamentos florestais de valor especial;
3. Isolamento de potenciais focos de ignição de incêndios;

As intervenções feitas nestas faixas são classificadas da seguinte forma: **Faixas de Redução de Combustível (FRC)** e **Faixas de Interrupção de Combustível (FIC)**. As FRC caracterizam-se pela remoção da vegetação de superfície e pelo corte de árvores de forma a criar uma separação entre copas, suprimindo também a parte inferior das mesmas (Figura 2.1). As FIC consistem na remoção total de todo o tipo de combustível dentro da faixa [10] (Figura 2.2).

No que diz respeito às FRC a distribuição da vegetação tem de obedecer às regras que estão definidas na lei [3, 23].

As regras estabelecem que a separação entre copas de árvores não pode ser inferior a 4 metros, sendo que entre eucaliptos ou pinheiros esta distância passa para 10 metros.



Figura 2.1: Exemplo de uma FRC na Nazaré. Figura 2.2: Exemplo de uma FIC em Silves.

A parte inferior das copas tem de ser suprimida até metade da altura da árvore, caso a árvore seja inferior a 8 metros. Árvores de tamanho superior deve ser realizada uma desarmação até os 4 metros a contar desde o solo.

No caso de combustíveis de superfície, a altura não pode exceder 50 cm no caso de arbustos. Já as ervas a altura máxima é de 20 cm [23]. No entanto, a quantidade deste tipo de combustíveis, dentro das faixa, não pode exceder os 2000 m<sup>3</sup>/ha.

### 2.2.2 Rede Primária de Faixas de Gestão de combustíveis (RPFGC)

A RPFGC é uma rede que cobre uma grande parte do território português, dividindo-o em parcelas que devem ter entre 500 ha e 10000 ha.

Esta rede é constituída por faixas de 125 metros de largura, dos quais 5 metros são relativos à via, 20 metros ( 10 em cada lado, adjacentes à via) são FIC e os restantes 100 metros ( 50 de cada lado) são FRC. A gestão da vegetação nestas faixas tem de respeitar as condições descritas na secção 2.2.1.

A RPFGC tem com principal objetivo cumprir a função 1 descrita anteriormente. Além desta função, estas faixas também são um elemento auxiliar na estruturação do combate aos incêndios, uma vez que as FGCI desta rede são construídas em zonas estratégicas.

### 2.2.3 Rede Secundária e Rede Terciária

As faixas das redes secundária e terciária são apoiadas nas redes viárias e florestais e ao redor de habitações e aglomeradas habitacionais. Estas redes são definidas pelos municípios e fazem parte do Plano Municipal de Defesa da Floresta Contra os Incêndios (PMDFCI).

As faixas criadas ao longo das redes viárias e florestais têm um perímetro mínimo de 10 metros em cada lado da via e são do tipo FIC. O mesmo também se aplica à rede ferroviária.

Faixas ao redor de habitações têm um raio de 50 metros e quando se tratam de aglomerados habitacionais esta distância passa a ser de 100 metros. A disposição da vegetação



Figura 2.3: Faixa da Rede Primária na zona de Silves.

nestas faixas é semelhante à referida na secção 2.2.1 com a adição de que a copa das árvores têm de estar a mais de 5 metros das habitações.

Linhas de distribuição elétrica de alta e média tensão também fazem parte destas redes. No caso de linhas de média tensão tem de ser criada uma faixa do tipo FIC com pelo menos 7 metros para cada lado do centro da linha. As linhas de alta tensão têm uma FIC de largura igual à envergadura das linhas e com mais 10 metros para cada um dos lados.

No caso de linhas de transporte de gás natural (gasodutos) tem de ser feita uma gestão de combustível numa faixa de largura igual ao superior a 7 metros para cada um dos lados, contados a partir o centro da conduta.

Estas faixas estabelecem perímetros que permitem proteger bens e infraestruturas, assim como as populações cumprindo as funções 2 e 3. Nas Figuras 2.4, 2.5 e 2.6 estão alguns exemplos de faixas destas redes.



Figura 2.4: Linha de alta tensão.



Figura 2.5: FGCI ao longo de uma estrada.



Figura 2.6: Faixa no redor de uma habitação

Para que as FGCI cumpram as suas funções é preciso mantê-las limpas, respeitando as condições estabelecidas. A limpeza e manutenção das FGCI é da responsabilidade do município, dos proprietários dos terrenos e das entidades privadas, no caso de linhas elétricas (alta e média tensão) e gasodutos.

Quando existem terrenos de privados que fazem parte da FGCI os proprietários são os responsáveis pela limpeza. Caso estes não realizem a devida limpeza no prazo indicado o município passa a ser o responsável por efetua-la.

### 2.2.3.1 Aspetos a Considerar na Análise das FGCI

Ao avaliar cada tipo de faixa individualmente e as condições do terreno onde estas faixas podem estar inseridas, podemos verificar que existem alguns fatores que podem afetar a vegetação dentro destas FGCI.

No geral as FGCI têm o propósito de impedir a propagação do fogo e proteger infraestruturas, sendo que no caso das estradas e rede primária, podem servir de pontos estratégicos no combate aos incêndios. Porém, estas faixas podem estar inseridas em contextos que podem dificultar a análise da vegetação.

Um fator que é comum a todos os tipos de faixa e pode dificultar a análise da vegetação nas FGCI através de imagens satélite é a existência de campos agrícolas. Ao analisar uma imagem satélite de uma zona onde existe um campo agrícola que intersecta a faixa, esta pode aparentar não ter sido intervencionada, embora a cultura agrícola instalada possa não ser propícia à propagação de fogo.

Nas faixas em redor das habitações e em parte das faixas da RPFGC, em que não é obrigatório por lei a remoção de todas as árvores, o comportamento da vegetação também é distinto.

Por outro lado as faixas ao longo de linhas elétricas, além da possibilidade de impedir a propagação do fogo, têm como principal função prevenir que ocorram descargas elétricas das linhas para árvores, que podem originar incêndios. A elevação do terreno pode ser determinante na gestão da vegetação destas faixas, uma vez que em alguns vales a linha elétrica pode passar a várias dezenas de metros do solo, sendo que não existe a necessidade de remover a vegetação.

No momento da análise de resultados obtidos através do processamento de imagens satélite é necessário ter em conta todos estes fatores uma vez que podem gerar resultados não esperados.

### 2.2.4 Informação de Referência das FGCI

Toda a informação sobre as FGCI é disponibilizada em formato vetorial pelo ICNE.

São disponibilizados ficheiros *shapefile* das FGCI de todos os concelhos de Portugal continental. É da responsabilidade dos municípios a criação destes ficheiros vetoriais e de estabelecer quais as faixas que fazem parte da rede secundária e terciária.

A representação das faixas é feita através de polígonos, e a cada um deles estão associados vários atributos que contêm informação sobre a faixa em questão (Tabela 2.1). Na Figura 1.1 pode ser visto exemplos de polígonos que representam parte das FGCI, numa zona do concelho de Mação.

Tabela 2.1: Atributos associados a cada uma das faixas e a respetiva descrição [32].

Nome do Campo	Descrição
ID_R_FGC	Número natural que identifica a FGCI.
ID_S_FGC	Número decimal com algarismo das unidades a corresponder ao ID_R_FGC e o das décimas a um número único e irrepetível, de forma a identificar a secção constituinte de determinada FGCI.
DATA_ACCAO	Data em que foi efectuado o levantamento das características no terreno (não corresponde à data de intervenção).
COD_INE	Código de Referenciação Territorial referente a distrito, concelho e freguesia.
DESC_FGC	Número natural (1 a 14) que descreve o tipo de faixa (e.g. estradas, linhas de alta tensão, aglomerados habitacionais, etc).
TIPO_FGC	Tipo de intervenção nas FGCI ( FIC ou FRC).
OBJEC_FUNC	Objectivo/função das faixas combustível (secção 2.2.1).
AREA	Área de cada secção das faixas expressa em hectares (ha).
RESP_GC	Identificação do responsável pela gestão de combustível.
INTER_AAAA	Tipo de intervenção a realizar nas FGCI (e.g. desramações, fogo controlado, moto-manual, aplicação de fitocidas, sem intervenção, etc), onde AAAA indica o ano da intervenção.
EXEC_AAAA	Meio de execução das intervenções nas faixas (e.g. meios próprios da autarquia, sapadores florestais, prestadores de serviços, etc), onde AAAA indica o ano da intervenção.
FIN_AAAA	Meio de financiamento para execução das faixas (e.g. autarquia, fundo florestal, etc), onde AAAA indica o ano da intervenção.
FASE_AAAA	Fase do projecto das FGCI (e.g. marcação, validação, execução, etc), onde AAAA indica o ano da intervenção.
OBSERV	Observações relevantes e que complementem a informação presente nos campos anteriores.

Estes atributos fornecem informação que permite, entre outras coisas, identificar o tipo da faixa, tipo de intervenção ou o ano da intervenção. Este tipo de informação é importante pois permite avaliar individualmente diferentes tipos de faixa ou realizar operações faixa a faixa.

Apesar da forma de preenchimento de cada um destes atributos estar muito bem



definida no guia técnico do PMDFCI [32], existem alguns concelhos onde o preenchimento não foi efetuado de forma correta ou onde vários campos foram deixados em branco.

Além destas omissões no preenchimento também existem incorreções na criação dos polígonos que definem os limites das FGCI. Foi observado em alguns casos que os limites das FGCI foram mal definidos, resultando que infraestruturas como casas ou estradas pertencessem ao interior da FGCI, quando apenas a vegetação em redor é que deveria fazer parte da faixa.

Na Figura 2.7a encontra-se um exemplo de uma área com várias habitações e aglomerados habitacionais, numa zona no norte do distrito de Castelo Branco, e na Figura 2.7b a representação realizada das faixas em redor das habitações, que inclui erradamente as habitações como pertencentes à faixa e atribui uma largura de faixa de 50m a habitações isoladas e a aglomerados habitacionais, quando deveriam ser 100m no caso dos aglomerados.

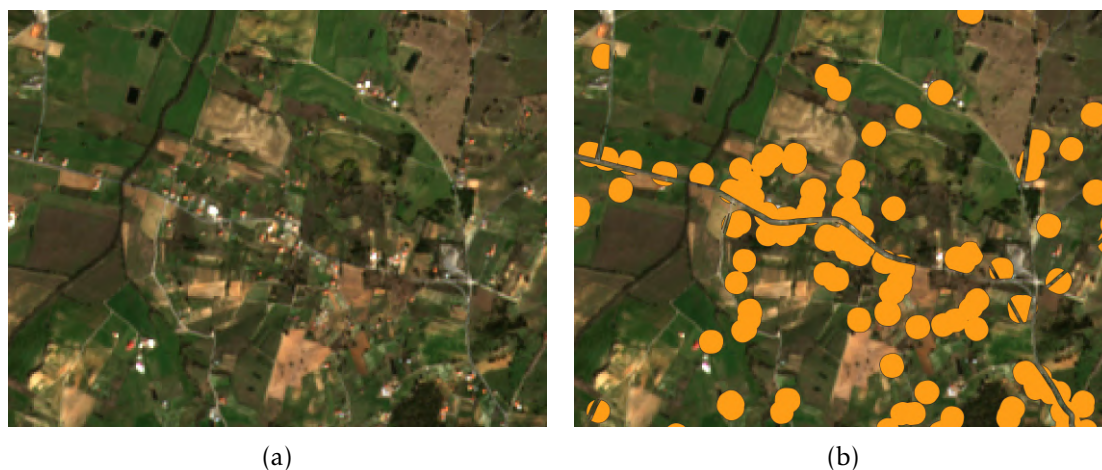


Figura 2.7: Exemplo de faixas mal definidas, na zona norte do distrito de Castelo Branco. (a) Imagem do satélite Sentinel-2 numa zona de habitações. (b) Polígonos representativos de faixas em redor a habitações, que foram mal definidos.

Contudo, no concelho que foi analisado, Mação, pôde ser verificado que o preenchimento dos atributos associados às faixas e a delineação dos limites das FGCI foram realizados corretamente, e não foram notadas irregularidades como as descritas anteriormente.

## 2.3 Detecção Remota

Existem vários satélites, que providenciam diferentes tipos de dados, e que são usados na área de deteção remota para realizar observações.

Satélites com sensores multiespectrais, que possuem bandas no infravermelho próximo e no infravermelho de onda curta, são recorrentemente usados. Também por vezes é usada a combinação destes satélites com satélites captam imagens de radar, obtendo assim um conjunto de diferentes características da área em estudo.

Apesar de existirem vários exemplos de satélites que fornecem estes dados, nesta dissertação foram usados os Sentinel-2 e Sentinel-1 pois disponibilizam publicamente os seus dados e têm boas resoluções espaciais.

### 2.3.1 Satélites Multiespectrais

Na área de detecção remota é recorrente o uso de satélites que possuem sensores multiespectrais. Existem vários satélites com este tipo de instrumentos que disponibilizam os seus dados publicamente, como é o exemplo dos Sentinel-2.

Os Sentinel-2A e 2B são satélites que fazem parte do programa *Copernicus*. Ambos têm um órbita síncrona em relação ao sol, passando no equador sempre à mesma hora, garantindo que o ângulo de incidência da luz solar é mantido, apenas variando dependendo da altura do ano.

O sensor multiespectral presente nestes satélites capta informação em 13 bandas, no visível no *near-infrared* (NIR) e no *short wave infrared* (SWIR) sendo que estas bandas têm resoluções espaciais diferentes: 10m, 20m e 60m. Na Tabela 2.2 encontram-se detalhadas as bandas disponibilizadas assim como os respetivos comprimentos de onda.

Tabela 2.2: Bandas disponibilizadas pelo Sentinel-2A (S2A) e Sentinel-2B (S2B) [42].

Banda	Resolução (m)	Comprimento de onda S2A (nm)	Comprimento de onda S2B (nm)
1 (Aerosol)	60	442.7	442.2
2 (Azul)	10	492.4	492.1
3 (Verde)	10	559.8	559.0
4 (Vermelho)	10	664.6	664.9
5 ( <i>red edge</i> - RE1)	20	704.1	703.8
6 ( <i>red edge</i> - RE2)	20	740.5	739.1
7 ( <i>red edge</i> - RE3)	20	782.8	779.7
8 (NIR)	10	832.8	832.9
8A (NIR_A)	20	864.7	864.0
9 (Vapor de água)	60	945.1	943.2
10 (SWIR)	60	1373.5	1376.9
11 (SWIR)	20	1613.7	1610.4
12 (SWIR)	20	2202.4	2185.7

Os dados recolhidos por estes satélites contêm imprecisões causadas por vários fatores e têm de ser processados antes de serem disponibilizados aos utilizadores, de maneira a corrigi-las. São fornecidos dois tipos de produtos, sendo que cada um deles é caracterizado pelo nível de processamento a que os dados foram sujeitos (Tabela 2.3).

Os produtos disponíveis podem ser usados para diferentes tipos de aplicações sendo que existem cuidados a ter no uso do produto 1C e algumas vantagens no uso do 2A que irão ser abordadas na secção 2.3.3.1.



Tabela 2.3: Descrição dos produtos do Sentinel-2 que são disponibilizados [41].

Nível do Produto	Descrição
1C	Disponibiliza imagem projetada em geometria cartográfica e com refletância no topo da atmosfera.
2A	Imagens com refletância na base da atmosfera geradas a partir o produto de nível 1C.

### 2.3.2 Synthetic-Aperture Radar (SAR)

Satélites do tipo SAR, como é caso dos Sentinel-1A e 1B, são usados com frequência em combinação com dados multiespectrais de outros satélites.

Este satélite tem três modos de operação, sendo que o principal modo, quando este satélite se encontra sobre terreno, é o *Interferometric Wide* (IW) que recolhe dados com uma resolução espacial de 5m x 20 m (uma única observação). Este modo oferece dados em polarização VV (*Vertical transmit – Vertical receive*) ou HH (*Horizontal transmit – Horizontal receive*), ou então em pares VV+HV ou VV+VH. Com o modo IW são oferecidos vários produtos: dados em bruto, *Single Look Complex* (SLC) e *Ground Range Detected* (GRD) com imagens de alta e média resolução. As imagens de maior resolução são oferecidas pelo produto GRD e têm uma resolução espacial de 20m x 22m, com uma amostragem de 10 metros [40].

A vantagem do uso deste tipo de satélites é que os comprimentos de onda do SAR permitem que o sinal penetre as nuvens, fazendo com que as condições atmosféricas não afetem o sinal, permitindo também realizar observações tanto de dia como de noite.

O instrumento SAR do Sentinel-1 tem uma banda do tipo C. Este tipo de banda, que devido ao seu comprimento de onda, reflete na camada superior das copas, faz com que os dados obtidos pelo Sentinel-1 tenham potencial para estimar algumas características da vegetação.

Apesar de ser mais frequente o uso de satélites com bandas multiespectrais, existem trabalhos que combinam os dados destes tipos de satélites, e que mostraram que esta combinação pode gerar bons resultados na estimativa de características da vegetação [7, 46].

### 2.3.3 Pré-processamento de Imagens Satélite

As imagens provenientes de satélites têm propriedades diferentes que variam de satélite para satélite, ou até mesmo de órbita para órbita. Desta forma é necessário que exista um cuidado no processamento destas imagens.

Estas situações podem resultar em erros de classificação ou em modelos imprecisos, assim para minimizar possíveis erros é fundamental realizar um pré-processamento das imagens.

### 2.3.3.1 Sentinel-2

Nos satélites multi-espectrais, como é o caso do Sentinel-2, podem ocorrer alguns tipos de anomalias quando uma imagem é captada, desde falhas instrumentais a erros de processamento ou interferências atmosféricas. Um exemplo comum é a ocorrência de nuvens que por vezes podem cobrir na totalidade da imagem captada.

A composição da atmosfera também pode afetar ligeiramente o sinal recebido, fazendo com imagens tiradas em momentos ou lugares diferentes sejam afetadas de diferentes formas. Como consequência resultados conseguidos a partir de imagens obtidas nestas condições podem não ser comparáveis [2].

Outra situação que pode ocorrer é duas imagens de uma determinada zona, tiradas por um satélite, mas em órbitas diferentes, podem conter erros de georreferenciação, fazendo com um mesmo ponto esteja em localizações diferentes nas duas imagens.

#### **Correções atmosféricas**

Alguns níveis de processamento dos produtos satélites Sentinel-2 apenas disponibilizam valores da refletância no topo da atmosfera, como é o caso do nível 1C (Tabela 2.3). Isto pode ser problemático quando estamos a lidar com dados multi-temporais uma vez que a composição da camada atmosférica não é constante, podendo variar dependendo da altura do ano ou até mesmo da localização.

Esta instabilidade tem impacto nos valores obtidos pelas bandas dos satélites e por consequência nos índices de vegetação calculados a partir delas, especialmente quando dados multi-temporais estão a ser usados [2, 18]. Assim é necessário realizar uma correção atmosférica para poder obter a refletância na base da atmosfera (ou superfície terrestre) e ser possível comparar valores obtidos em datas diferentes.

Existem ferramentas e técnicas que transformam os produtos com dados de refletância no topo da atmosfera em refletância à base da atmosfera. Para o Sentinel-2 é disponibilizado um produto do nível 2A (com correção atmosférica aplicada), e também é disponibilizada a ferramenta *Sen2Cor* [39], que gera um produto do nível 2A a partir de um produto de nível 1C.

#### **Correções de georreferenciação**

O pré-processamento realizado às imagens satélite, antes de serem disponibilizadas aos utilizadores, usa pontos de controlo terrestres para georreferenciá-las. Porém existe um erro associado à precisão da georreferenciação [47].

Ter dados precisos sobre a georreferenciação é extremamente importante uma vez que, no contexto desta dissertação, estão a ser analisadas FGCI que podem ter apenas 10 metros de largura, o que equivale a um píxel em algumas bandas do Sentinel-2.

Atualmente a performance no registo geográfico de imagens multi-temporais do Sentinel-2 é de cerca de 11 metros, com um nível de confiança de 95% [36]. Contudo, foi criada uma imagem de referência a nível global (GRI) que irá servir de referência para

aperfeiçoar as geometrias do terreno, aumentando assim a performance de 11 metros para cerca de 3 metros [11] e que neste momento está em fase de validação [36].

### 2.3.3.2 Sentinel-1

Tal como nos satélites multi-espectrais, também nos satélite do tipo SAR existem alguns fatores que podem comprometer a qualidade dos dados captados pelo satélite.

Um dos fatores a ter em conta é a radiação térmica emitida pelo solo que pode introduzir ruído na imagem captada. Outro fator que é necessário considerar é o da órbita do satélite que pode variar ligeiramente, sendo necessário corrigir usando informação precisa da órbita que foi realmente realizada.

Através da aplicação *Sentinel Application Platform* (SNAP), disponibilizada pela *European Space Agency* (ESA), podem ser processadas as imagens do Sentinel-1 e minimizar o impacto destes problemas nas imagens.

## 2.4 Índices de Vegetação

Na literatura são usados diversos índices que permitem estimar várias características da vegetação como por exemplo: o seu vigor, densidade, conteúdo de água, etc. Com recurso às bandas disponibilizadas pelos satélites é possível calcular alguns índices que estimam estas características.

### 2.4.1 Normalized Difference Vegetation Index

Um índice bastante usado em aplicações na área de deteção remota, é o *Normalized Difference Vegetation Index* (NDVI), que é calculado usando a banda no vermelho (RED) e no infravermelho próximo (NIR) através da equação 2.1.

$$NDVI = \frac{NIR - RED}{NIR + RED} \quad (2.1)$$

Os valores deste índice variam entre -1 e 1, quando os valores da banda vermelha aproximam-se de 0 o NDVI aproxima-se de 1, o que indica a existência de vegetação com bastante vigor ou densa. Quando os valores do infravermelho próximo estão próximos de 0 o NDVI tende para -1, enquanto que valores próximos ou inferiores a zero indicam vegetação seca, inexistência de vegetação ou água.

Na Figura 2.8b podemos ver um exemplo do NDVI numa zona do distrito de Santarém, em que valores abaixo do 0,35 mostram zonas de vegetação seca ou de habitação e valores a cima de 0,65 zonas vegetação com mais vigor ou mais densa.

O NDVI tem mostrado ao longo do tempo obter bons resultados na classificação e na estimativa de características biofísicas da vegetação [5, 26, 29]. Pelletier et al. [29] mostrou que a combinação do NDVI com as várias bandas disponibilizadas por satélites como o

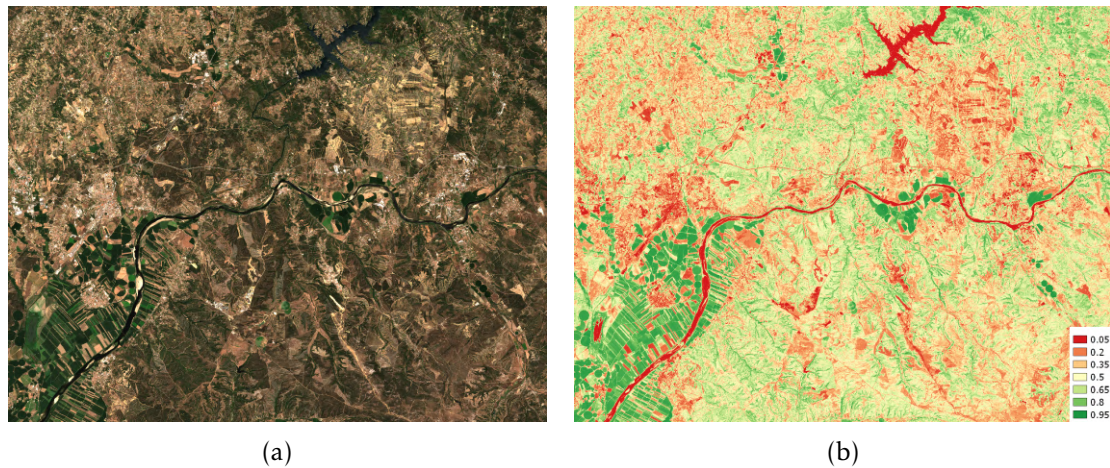


Figura 2.8: (a) Imagem a cores de um zona do distrito de Santarém. (b) Resultado do cálculo do NDVI para a mesma zona.

Landsat 8, gera bons resultados na classificação da cobertura do terreno. Müller et al. [26] produziu séries temporais usando algumas métricas (média, mediana, desvio padrão, etc) de bandas espectrais e do NDVI para classificar tipos de vegetação, obtendo uma exatidão global de 93%.

#### 2.4.2 Soil-Adjusted Vegetation Index

Outro índice que também ocorre com alguma frequência na literatura é o *Soil-Adjusted Vegetation Index* (SAVI), que foi proposto por Huete (1988) [20]. Este índice é uma variante do NDVI, e como tal a sua equação 2.2 usa as mesmas bandas, mas adicionando um fator  $L$  para tentar atenuar o impacto que o solo tem nos valores da refletância.

$$SAVI = (1 + L) * \frac{NIR - RED}{NIR + RED + L} \quad (2.2)$$

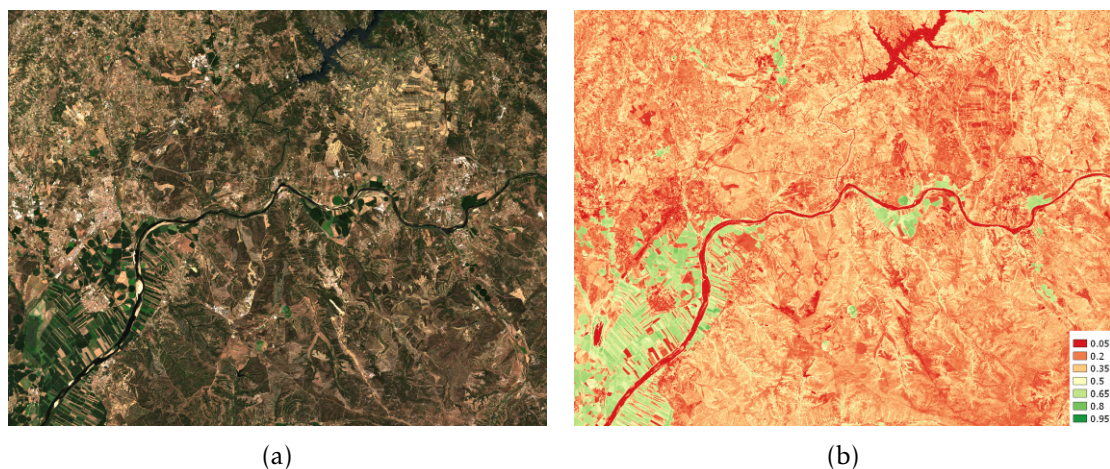


Figura 2.9: (a) Imagem a cores de um zona do distrito de Santarém (b) Resultado do cálculo do SAVI para a mesma zona.



Tal como o NDVI os valores deste índice variam entre -1 e 1 e o valor do fator  $L$  varia entre 0 e 1, sendo que foi usado o valor empírico de  $L = 0,5$ . Quando usado  $L = 0$  a expressão é igual à do NDVI 2.1 sendo mais indicado para áreas com muita densidade de vegetação, quando o valor de  $L$  aproxima-se de 1 este índice é mais apropriado para áreas com baixa densidade de vegetação [31].

Alguns trabalhos usam este índice em junção com outros na estimativa de biomassa e na classificação da cobertura do terreno [31, 44, 46]. Apesar deste índice não se destacar de outros, em nenhum destes trabalhos em específico, encontra-se entre os melhores na estimativa de biomassa [46].

### 2.4.3 Simple Ratio

O índice de vegetação *Simple Ratio* (SR) é calculado usando a expressão 2.3, que utiliza as bandas NIR e a banda no vermelho. O valor mínimo que este índice pode ter é 0 e não existe um valor máximo associado. Valores próximos de 0 sugerem a existência de solo exposto, vegetação seca ou água, enquanto que valores mais elevados indicam a presença de vegetação.

$$SR = \frac{NIR}{RED} \quad (2.3)$$

Na Figura 2.10 está um exemplo deste índice onde podemos observar que valores entre 10 e 20 mostram a existência de vegetação, sendo que os valores mais elevados correspondem a zonas de campos agrícolas.

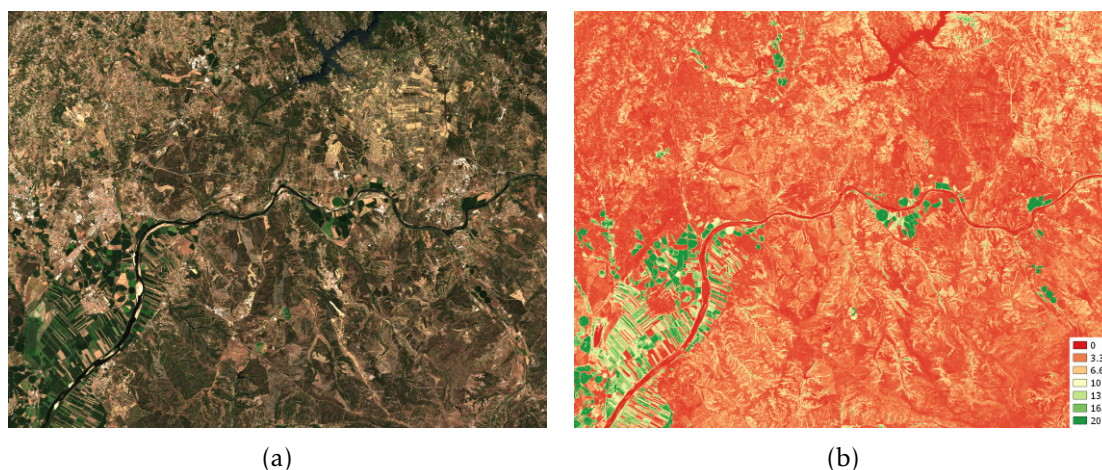


Figura 2.10: (a) Imagem a cores de um zona do distrito de Santarém (b) Resultado do cálculo do SR para a mesma zona.

Vafaei et al. [46] usou alguns índices de vegetação para estimar a biomassa. O SR foi um dos três índices que se destacaram, mostrando ter uma correlação com os níveis de biomassa presentes no terreno.

#### 2.4.4 Inverted Red-Edge Chlorophyll Index

Frampton et al. [14] apresentaram o índice *Inverted Red-Edge Chlorophyll Index* (IRECI) que usa as bandas *red-edge* do Sentinel-2 e que mostrou ser um bom indicador de características biofísicas da vegetação. Este índice é calculado usando a equação 2.4 que usa a banda no vermelho e as bandas no *red edge*: banda 5 (RE1), 6 (RE2) e 7 (RE3).

$$IRECI = \frac{RE3 - RED}{\frac{RE1}{RE2}} \quad (2.4)$$

Na Figura 2.11b podemos ver um exemplo deste índice, em que valores próximos de 0 indicam zonas sem vegetação, com vegetação seca ou com água, enquanto que valores próximos ou superiores a 1 indicam a presença de vegetação. Quanto mais elevado o valor deste índice maior a densidade ou vigor da vegetação.

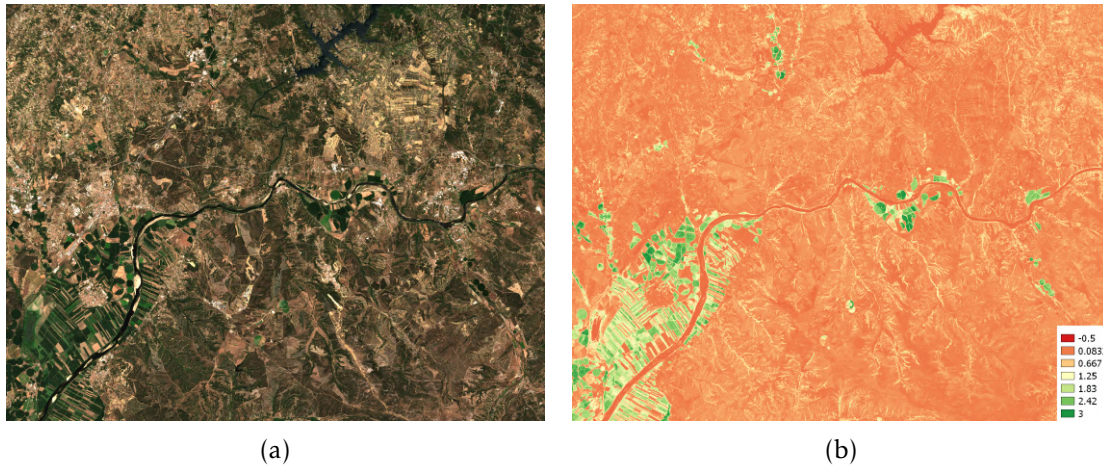


Figura 2.11: (a) Imagem a cores de um zona do distrito de Santarém (b) Resultado do cálculo do IRECI para a mesma zona.

No trabalho apresentado este índice foi o melhor a estimar a cobertura da vegetação, tanto em áreas com baixa cobertura como em áreas de grande cobertura. O mesmo foi observado por Castillo et al. [7] na estimativa de biomassa, onde o IRECI em combinação com dados de elevação do terreno obteve o erro mais baixo a estimar a quantidade de biomassa presente no terreno.

#### 2.4.5 Normalized Difference Water Index

Existe também outro tipo de índices que tentam avaliar a quantidade de água presente na vegetação e que podem indicar se a vegetação se encontra seca ou não, um desses índices é o *Normalized Difference Water Index* (NDWI) [15].

Ao contrário dos índices anteriores, o NDWI usa apenas bandas no infravermelho (a banda NIR e a banda de infravermelho de onda curta (SWIR)) e é calculado através da expressão 2.5.

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR} \quad (2.5)$$

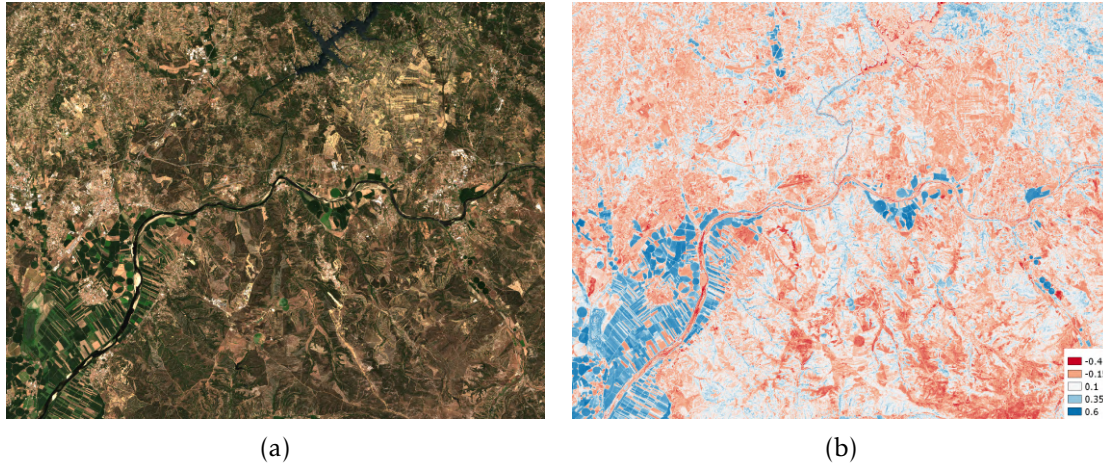


Figura 2.12: (a) Imagem a cores de um zona do distrito de Santarém (b) Resultado do cálculo do NDWI para a mesma zona.

Os valores deste índice variam entre -1 e 1, valores positivos indicam vegetação com maior conteúdo de água, enquanto que valores negativos ou próximos de 0 indicam vegetação seca ou inexistência de vegetação. Na Figura 2.12b podemos observar uma imagem de cor falsa do NDWI, onde os tons azul indicam zonas onde a vegetação tem um maior conteúdo de água (neste exemplo correspondem a áreas agrícolas) e tons de vermelho indicam vegetação seca ou inexistência de vegetação.

Inglada et al. [22] usou o NDWI e o NDVI juntamente com valores de refletância para classificar a cobertura do terreno em diferentes países, obtendo na maioria dos casos valores de exatidão superiores a 85%.

## 2.5 Trabalhos Relacionados

Na área de detecção remota existem variados trabalhos relacionados com a classificação de áreas florestais, urbanas ou de culturas agrícolas com recurso a imagens de satélite.

O lançamento dos satélites Sentinel-2A e Sentinel-2B trouxe novas oportunidades de desenvolvimento de aplicações em diversas áreas de estudo. Trabalhos realizados com o objetivo de identificar a cobertura do terreno, classificando espécies de árvores e tipos de plantações agrícolas, com recurso a dados de satélites, mostram o potencial que as bandas infravermelhas destes satélites têm para este tipo de aplicação [6, 21, 29].

O uso de satélites do tipo SAR tem a possibilidade de ser vantajoso para este tipo de aplicações uma vez que obtém dados da superfície terrestre independentemente das condições atmosféricas ou da altura do dia. Existem trabalhos que combinam os dados destes dois tipos de satélites para estimar classificar ou estimar características da vegetação e esta combinação tem mostrado bons resultados.



No contexto desta dissertação é importante perceber em que estado se encontra a vegetação dentro das FGCI. Assim, aspetos como o vigor e cobertura da vegetação podem ser indicadores de que uma faixa foi intervencionada ou está limpa.

Assim, nesta secção são abordados alguns trabalhos que têm como tema a monitorização e classificação de vegetação e que serviram de base de conhecimento na definição da abordagem usada nesta dissertação.

### 2.5.1 Classificação de Vegetação

Existem vários trabalhos que abordam a classificação de vegetação usando diferentes satélites. São usados com mais frequência satélites multiespectrais (e.g. Sentinel-2, Landsat-8, MODIS, etc), embora outros trabalhos combinam este tipo de satélites com satélites do tipo SAR (e.g. Sentinel-1) [12], sendo que as temáticas mais recorrentes são a monitorização de florestas, o mapeamento da cobertura do terreno e a agricultura [49].

Podemos encontrar trabalhos em que o foco é classificar tipos de culturas [6, 21, 30], outros que classificam espécies de árvores em zonas florestais [21] e ainda uma classificação mais abrangente que se foca em distinguir diferentes tipos de cobertura do terreno (e.g. zonas florestais, urbanas, agrícolas, etc) [12, 26, 27].

Existem trabalhos que apenas recorrem aos dados das bandas dos satélites usados [21, 27], embora na maioria dos casos sejam usados índices de vegetação como indicadores de características da vegetação.

Alguns trabalhos recorrem a séries temporais de forma a extrair mais informação sobre a vegetação [6, 26, 29]. O uso destas séries já revelou obter melhores resultados na classificação de vegetação e cobertura do terreno do que o uso de dados referentes a apenas uma data [17, 26]. Ainda assim, dados de apenas uma data já revelaram gerar bons resultados na cobertura do terreno, usando vários algoritmos de aprendizagem automática [27].

Immitzer et al. [21] classificou a vegetação numa zona agrícola e numa zona florestal usando o algoritmo *Random Forest* (RF). Na área agrícola foram classificadas 8 tipos de culturas e na área florestal 6 espécies de árvores. Foram comparadas duas abordagens: baseada em píxeis e baseada em objetos, e foi observado que na área florestal a abordagem baseada em objetos obteve melhores resultados, no entanto na área agrícola a abordagem baseada em píxeis foi a melhor.

Embora o método usado com mais frequência em trabalhos na área de deteção remota seja recorrendo ao uso de algoritmos de aprendizagem automática, existem outros métodos que permitem comparar semelhanças entre séries temporais e realizar estimativas. Belgiu et al. [6] comparou o algoritmo *Time-Weighted Dynamic Time Warping* (TWDTW) com o algoritmo RF na classificação de tipos de culturas usando séries temporais do NDVI, em três áreas de teste, sendo que em duas dessas áreas o TWDTW obteve melhores resultados que o RF. Foram ainda comparadas as abordagens baseada em objetos e em píxeis e em ambos os algoritmos as duas abordagens tiveram resultados muito próximos.



### 2.5.2 Monitorização de Culturas

Outro tipo de aplicações que recorre a imagens satélite é a monitorização de culturas. Com a lançamento de satélites de maior resolução, como é o caso de Sentinel-2 é possível realizar observações mais detalhadas de campos agrícolas e monitorizar as culturas ao longo do ano.

Alguns trabalhos focam-se na monitorização de diferentes culturas, recorrendo ao índices de vegetação, de forma a perceber o impacto que fatores naturais ou externos têm nas culturas [9, 35, 43].

Clevers et al. [9] usou dados do Sentinel-2 e índices de vegetação para analisar zonas de 30m x 30m de um campo de batatas, de forma a perceber a cobertura da plantação, em cada uma das zonas, e comparar o efeito de diferentes tipos de fertilizante. Também usando o mesmo tipo de dados Rozenstein et al. [35] estimou o consumo de água de plantações de algodão, comparando com medições realizadas no terreno, e observou que existe uma forte correlação ( $R^2 = 0.94$ ) entre o NDVI e o consumo de água desta planta.

Outros exemplos de aplicações relacionadas com culturas tentam estimar a produção de uma determinada cultura recorrendo indicadores calculados a partir de medições no terreno, como o *Leaf Area Index* (LAI), e índices de vegetação [38, 43]. Setiyono et al. [43] usou dados SAR do Sentinel-1 e dados multiespectrais do satélite MODIS para estimar a produção de arroz, comparando com dados obtidos no terreno, obtendo exatidões superiores a 10%.

### 2.5.3 Outros Trabalhos

Alguns trabalhos na literatura tentam estimar a quantidade de biomassa presente na vegetação, mostrando existir uma correlação entre os níveis de biomassa e alguns índices de vegetação. O uso destes índices em combinação com as bandas multiespectrais dos satélites e com dados de satélites do tipo SAR [7, 44, 46] têm mostrado ser potenciais indicadores dos níveis de biomassa.

Castillo et al. usou dados o Sentinel-2, Sentinel-1 e dados de elevação do terreno, juntamente com alguns índices de vegetação para estimar a quantidade de biomassa presente em florestas de mangue nas Filipinas. Os resultados mostraram existir uma correlação dos níveis de biomassa com alguns índices de vegetação e o LAI.

Embora existam diversos trabalhos que abordam a monitorização e classificação de vegetação recorrendo à deteção remota, não existem muitos exemplos que abordem a temática das FGCI. Maltsev et al. [13] usou o NDVI, juntamente com informação geográfica de estradas e campos agrícolas, para identificar FGCI através de solo descoberto. Num contexto diferente Sarantis-Angelos G. Liampas et al. [37] comparou a exatidão de três modelos digitais de superfície, para estimar a elevação do terreno nas áreas que dizem respeito às FGCI, numa zona norte da Grécia.

Contudo, nenhum destes exemplos analisa o estado da vegetação dentro destas faixas.

### 2.5.4 Pontos Essenciais

Após a leitura de vários trabalhos da área de detecção remota foi possível perceber os aspectos que envolvem o processamento e análise de dados de imagens satélite, no que diz respeito à monitorização e classificação de diferentes tipos de vegetação.

A combinação da informação disponibilizada por diferentes tipos de satélite tende a melhorar a exatidão dos resultados. Em específico, a combinação de dados multiespectrais com dados de polarização mostraram bons resultados, pois em conjunto oferecem um maior número de características da área que está a ser analisada.

Além dos diferentes tipos de satélites, os índices de vegetação calculados recorrendo às bandas multiespectrais dos satélites, têm um papel fundamental uma vez que permitem estimar características da vegetação. O índice a destacar é o NDVI, que é usado com alguma frequência e tem mostrado bom desempenho em diferentes tipos de aplicações.

Embora não tenham sido encontrados trabalhos que tratem o problema específico do estado da vegetação dentro das FGCI, existem alguns trabalhos em que a dimensão das áreas que foram analisadas podem assimilar-se com as dimensões de algumas FGCI [9, 31]. Apesar das áreas serem de reduzida dimensão e pouco maiores que a resolução do satélite os valores do NDVI obtidos são semelhantes aos valores de imagens de maior resolução [31].

Toda a informação apresentada nestes trabalhos serve de base de conhecimento para estabelecer uma estratégia adequada no sentido de efetuar a classificação de intervenções e monitorização do estado da vegetação nas FGCI.

## 2.6 Conclusões

Os satélites disponibilizados pelo programa *Copernicus* e da informação vetorial das FGCI disponibilizada pelo o ICNF fornecem uma base de informação que permite criar uma abordagem para estimar o estado da vegetação dentro das FGCI.

As FGCI apresentam diferentes características que variam consoante o tipo de faixa, e como tal é preciso ter presentes as especificidades de cada um dos tipos de faixa de forma a definir uma abordagem apropriada e posteriormente realizar uma análise correta dos dados.

Alguns aspetos técnicos dos satélites podem ter influência nos resultados analisados. É necessário ter um cuidado especial quando séries temporais estão a ser analisadas pois existem múltiplos fatores que podem influenciar os dados das imagens dos satélites. Como tal, é fundamental realizar um pré-processamento das imagens de forma a corrigir potenciais erros que possam colocar em causa a validade dos resultados.

Os índices de vegetação podem ser utilizados ou combinados de forma a obter uma estimativa mais precisa das características da vegetação. Estes índices em combinação com as bandas dos satélites Sentinel-1 e Sentinel-2, fornecem um conjunto de características que permite estimar as condições da vegetação.

Existem inúmeros trabalhos relacionados com a monitorização e classificação de áreas de vegetação que recorrem a imagens de satélites multiespectrais e SAR [7, 12, 46], sendo que muitos deles recorrem a índices de vegetação para analisar diferentes características da vegetação. Embora estes trabalhos não tratem do caso específico da vegetação nas FGCI, fornecem uma base de informação útil para definir a abordagem a usar neste trabalho.

Todos estes aspetos são importantes e são a base da abordagem a usar para determinar o estado de limpeza ou de intervenção das FGCI.



## CLASSIFICAÇÃO AUTOMÁTICA EM DETECÇÃO REMOTA

### 3.1 Introdução

Na área de detecção remota a maior parte das vezes o objeto em estudo é a vegetação. Existem várias abordagens usadas de forma a monitorizar ou classificar vegetação em diversos contextos.

A técnica mais usada para realizar esta análise é recorrendo a algoritmos de aprendizagem automática. O uso destes algoritmos em junção com a informação disponibilizada por satélites, torna possível fazer previsões ou classificações muitas das vezes com bons resultados.

Existem abordagens diferentes para diferentes tipos de problemas, assim como estratégias de forma a melhorar a performance dos algoritmos, uma vez que o processamento de imagens satélite envolve a manipulação de grandes quantidades de dados.

Assim, neste capítulo são descritos alguns dos algoritmos de classificação, abordagens de classificação usadas na literatura, e métodos para melhorar tempos de execução recorrendo à seleção de características. Finalmente são apresentadas algumas métricas usadas para validar os resultados destes algoritmos.

### 3.2 Algoritmos de classificação

Na área de detecção remota o recurso a algoritmos de aprendizagem automática é bastante comum [21, 24, 27, 33]. Existem algoritmos que são referidos com alguma frequência na literatura, alguns deles são: *Random Forest* (RF), *Support Vector Machines* (SVM) e *K-Nearest Neighbors* (KNN).

Existem vários tipos de abordagens e estratégias que, recorrendo a estes algoritmos, tentam melhorar a precisão dos modelos. Na literatura existem variados exemplos de combinação de algoritmos com índices de vegetação e com bandas de diferentes satélites para tentar melhorar a precisão dos modelos em aplicações relacionadas com vegetação.

Nesta secção irão ser abordados alguns dos algoritmos mais recorrentes, assim como um algoritmo mais recente o *Extreme Gradient Boosting* (XGBoost), e o contexto no qual estes foram usados.

### 3.2.1 *K-Nearest Neighbors*

O *k-nearest neighbors* (KNN) é um algoritmo não paramétrico que atribui a classe mais comum entre os  $K$  pontos mais próximos, ao ponto que está a ser classificado. O cálculo da distância para obter os pontos mais próximos pode ser feito através da distância euclidiana ou de *Manhattan*. Este algoritmo por vezes é usado em aplicações de detecção remota e tem mostrado ter uma boa capacidade de classificação com recurso a imagens multiespectrais [19].

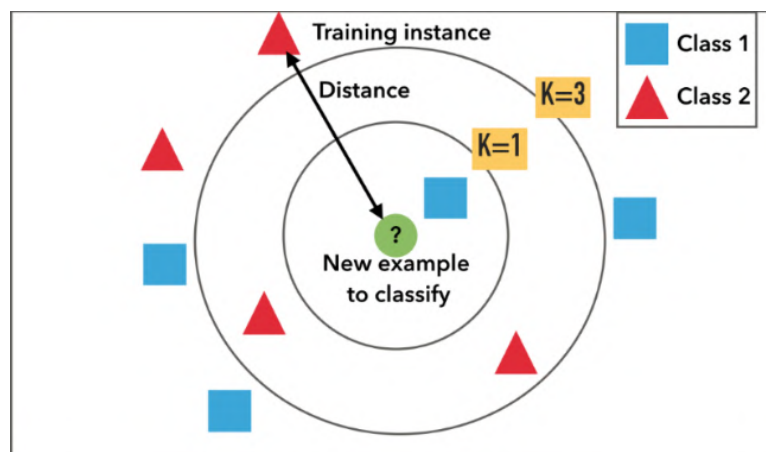


Figura 3.1: Exemplo do KNN com valores de  $K=1$  e  $K=3$ . (fonte: <https://medium.com/datadriveninvestor/knn-algorithm-and-implementation-from-scratch-b9f9b739c28f>)

Noi et al. [27] comparou alguns algoritmos na classificação da cobertura do terreno e detetou que o KNN é menos sensível com conjuntos de treino não balanceados, embora isso varie com o tamanho dos conjuntos de treino, sendo que com conjuntos de treino maiores a sensibilidade é menor. Também foi observado que em todos os testes realizados com este algoritmo o menor erro foi obtido quando foi considerando apenas o vizinho mais próximo.

### 3.2.2 *Support Vector Machine*

O *Support Vector Machine* (SVM) é um algoritmo de aprendizagem automática que tem como objetivo encontrar um hiperplano que realiza uma separação linear entre dois

conjuntos de pontos.

Podem existir vários hiperplanos que realizam uma separação entre dois conjuntos de pontos, contudo o hiperplano ótimo é o que tem a margem máxima entre ele e os pontos das duas classes (Figura 3.2).

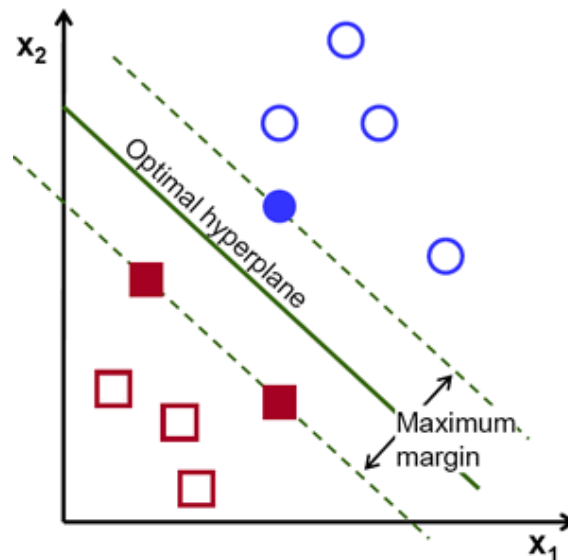


Figura 3.2: Exemplo de uma separação linear do SVM. (fonte: [https://docs.opencv.org/2.4/doc/tutorials/ml/introduction\\_to\\_svm/introduction\\_to\\_svm.html?highlight=svm](https://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html?highlight=svm))

Também é possível realizar separações não lineares recorrendo a funções *kernel*. Estas funções mapeiam um conjunto de pontos para um espaço dimensional superior, onde podem ser separados através de um hiperplano. Uma função de *kernel* bastante usada é o *Radial Basis Function* (RBF) e tem mostrado obter bons resultados em trabalhos de detecção remota, relacionados com a classificação de vegetação e cobertura do terreno [27, 29, 33].

Qian et al. [33] comparou alguns algoritmos de aprendizagem automática na classificação da cobertura do terreno, e observou que o SVM foi algoritmo que obteve os melhores resultados na grande maioria dos testes realizados.

### 3.2.3 Random Forest

Um dos algoritmos mais recorrentes na literatura relacionada com detecção remota é *Random Forest* (RF). Este algoritmo usa árvores de decisão e *bootstrap aggregating*, também conhecido por *bagging*, que consiste na criação de vários subconjuntos de dados, onde os elementos são escolhidos aleatoriamente e com repetição, para treinar as árvores de decisão e realizar classificações ou regressões.

Cada árvore de decisão é construída usando conjuntos de dados diferentes, gerados a partir do conjunto de treino usando *bootstrap*, e usando conjuntos de características

diferentes. Cada uma das árvores gera uma classificação, sendo que a classificação final é classe mais comum entre todas as árvores de decisão.

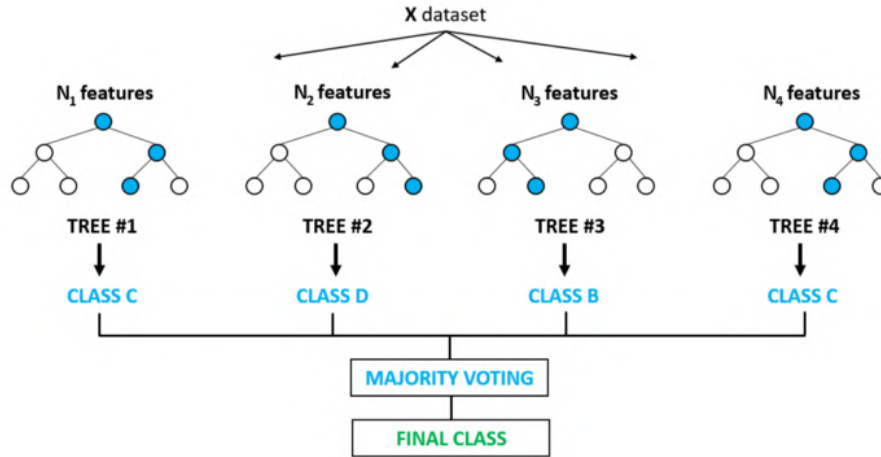


Figura 3.3: Exemplo do RF e das suas árvores de decisão. (fonte: <https://medium.com/@ar.ingenious/applying-random-forest-classification-machine-learning-algorithm-from-scratch-with-real-24ff198a1c57>)

Alguns trabalhos apenas recorrem a este algoritmo para realizar classificação de culturas e da cobertura do terreno [6, 21], por já ter mostrado bons resultados, que variam pouco, independentemente dos valores dos parâmetros, e ao mesmo tempo sendo bastante eficiente [29].

### 3.2.4 Extreme Gradient Boosting

*Extreme Gradient Boosting* (XGBoost) [8] é uma nova implementação de uma *framework* de *gradient boosting*, e que pode ser usada para fazer regressões, classificação ou ranking.

Tal como no RF este algoritmo usa árvores de decisão, mas em vez de ser usado *Bootstrap-Aggregating* é usado *boosting*, isto é, as árvores são construídas sequencialmente, sendo atribuído um peso aos elementos que foram classificados incorretamente pela árvore anterior, para que esses elementos sejam escolhidos pela árvore seguinte e possivelmente classificados correctamente. A classificação final de um dado elemento é a classe mais comum entre todas as árvores.

Uma das vantagens do XGBoost é o facto deste poder ser executado em paralelo, podendo correr em CPU, GPU ou em *frameworks* de computação distribuída como o *Hadoop* ou *Spark*. Este algoritmo já mostrou ser robusto em diferentes tipos de aplicações, destacando-se muitas vezes como a melhor solução em diversas competições [8].

Jiang et al. [24] usou o XGBoost e o RF para mapear plantações de cana de açúcar no sul da China, usando imagens do Sentinel-1. Ambos os algoritmos obtiveram resultados semelhantes, embora o XGBoost tenha mostrado ser mais resistente a *overfitting* e notoriamente mais rápido que o RF.



### 3.3 Abordagens de classificação

Existem duas abordagens que, em detecção remota, costumam ser usadas para classificar vegetação ou culturas agrícolas: classificação baseada em píxeis e classificação baseada em objetos.

Quando está a ser usada classificação baseada em píxeis, cada píxel, e os dados que estão associados a ele, é considerado como um exemplo do conjunto de treino/teste, ignorando por completo a informação dos píxeis vizinhos.

Na classificação baseada em objetos é tida em conta a informação dos píxeis vizinhos. Se estes forem semelhantes de acordo com o que está a ser analisado, o conjunto dos píxeis semelhantes é considerado um objeto, em que todos os píxeis que pertencem a esse objeto têm as mesmas características.

Uma das vantagens da abordagem baseada em objetos é que reduz significativamente o conjunto de dados, o que para aplicações que usam uma grande quantidade de dados pode ser interessante a fim de melhorar a performance de execução.

Belgiu et al. [6] aplicou estas duas abordagens na classificação de tipos de plantações e obteve melhores resultados em todas as áreas analisadas usando a classificação baseada em objetos. Além de melhores resultados a nível de classificação, a performance computacional foi claramente superior à da classificação baseada em píxeis.

Immitzer et al. [21] também realizou uma comparação entre estas duas abordagens na classificação de culturas agrícolas e de espécies de árvores e a abordagem baseada em objetos obteve uma exatidão superior à baseada em píxeis na classificação de espécies de árvores, no entanto na classificação de culturas a abordagem baseada em píxeis foi a melhor.

Estes resultados mostram que a abordagem baseada em objetos pode ser vantajosa em aplicações relacionadas com classificação de vegetação, tendo a vantagem de ter tempos de execução inferiores à abordagem baseada em píxeis.

### 3.4 Seleção de Características

Existem várias características que podem ser associadas a um píxel das imagens satélite, sendo parte dessas características as diversas bandas do satélite e os vários índices de vegetação que podem ser calculados a partir delas. Ao juntarmos a isto o uso de imagens temporais, pode ser facilmente gerado um grande número de características.

Como tal pode ser benéfico identificar e selecionar as características mais importantes e que melhor identificam os dados ou remover características que estão correlacionadas. Na literatura existem alguns exemplos do uso de seleção de características que recorrem a diferentes algoritmos.

Um dos métodos que pode ser usado para realizar esta seleção é a regressão LASSO, e existem alguns exemplos da sua utilização em aplicações na área de detecção remota [45, 48]. Takayama et al. [45] usou o LASSO de forma a determinar quais as bandas

multiespectrais ótimas para estimar a quantidade de biomassa, e ao mesmo tempo reduzir o conjunto de características.

Existem outras técnicas como o *Principal Component Analysis* (PCA) ou que recorrem a *deep learning* que também já foram aplicadas em trabalhos de detecção remota e que mostraram obter bons resultados [34].

Em conjuntos de dados de grandes dimensões e com um elevado número de características, que podem fazer com que os tempos de processamento sejam elevados, o uso destes métodos pode ser benéfico.

### 3.5 Validação

É importante ter informação com a qual comparar os resultados gerados pelos modelos criados. Apesar de existir muita informação nos ficheiros vetoriais disponibilizados pelo ICNF, por vezes esta não se encontra correta ou atualizada.

Através de uma colaboração com a Câmara Municipal de Mação foi possível obter informação fidedigna sobre as FGCI que foram intervencionadas no ano 2018 e que foi utilizada para validar os modelos. Nesse ano foram intervencionadas faixas ao longo de estradas e linhas elétricas, e a informação foi disponibilizada em formato vetorial. Na Figura 3.4 podem ser vistas a branco algumas faixas que foram intervencionadas.

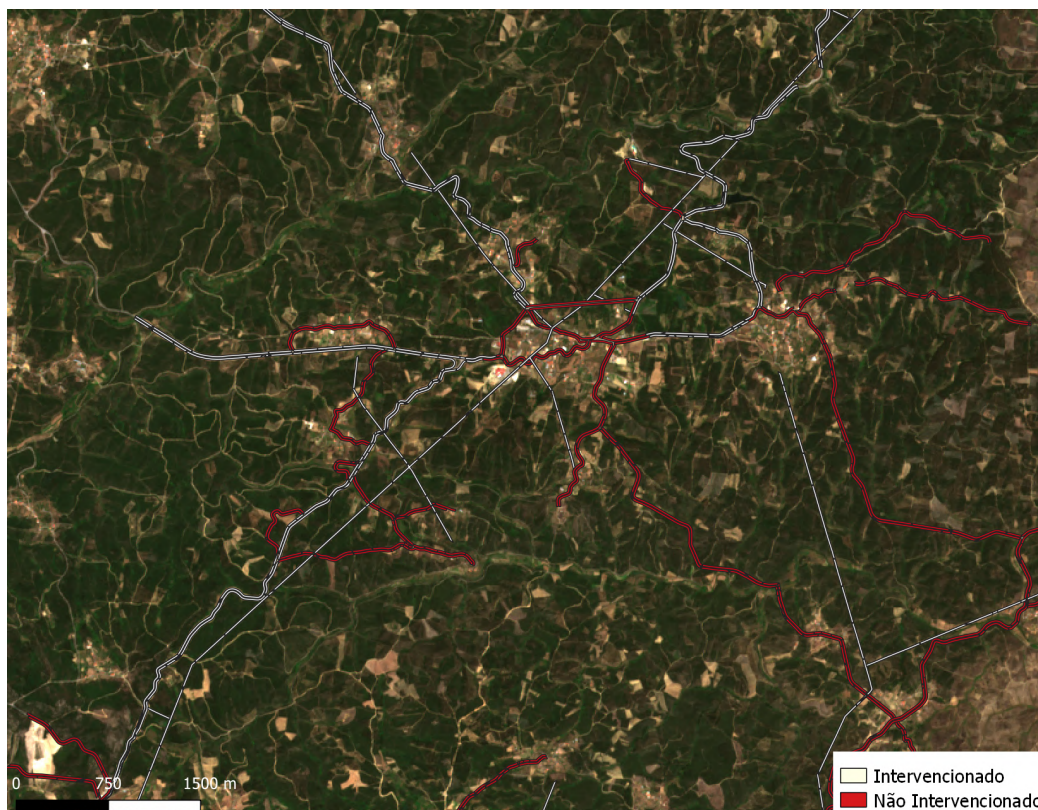


Figura 3.4: Faixas ao longo de estradas e de linhas elétricas, no norte de Mação, que foram intervencionadas no ano 2018.

Este tipo de informação é essencial para realizar a validação dos modelos criados. Existem várias métricas que permitem quantificar a qualidade dos modelos criados e que são usadas como referência para aperfeiçoar os modelos.

Métricas como a precisão ou o *recall* são uma boa base para avaliar a qualidade de um determinado modelo, mas existem outras métricas como o F1-score ou o Kappa fornecem mais informação sobre os resultados do modelo. Também podem ser usadas tabelas como as matrizes de confusão que, em junção com as métricas, permitem analisar mais detalhadamente os resultados obtidos.

### 3.5.1 Matriz de Confusão

Uma boa forma de analisar os resultados gerados por um classificador é através de uma matriz de confusão (Tabela 3.1).

As matrizes de confusão indicam os elementos que foram classificados de forma correta ou incorreta por um classificador, permitindo comparar com as classes de referência.

Através desta matriz podemos ver a distribuição de classes e também em que classes é que o classificador acertou. Além disso os valores presentes numa matriz de confusão permitem calcular várias métricas que possibilitam avaliar a qualidade do classificador.

Tabela 3.1: Exemplo da estrutura de uma matriz de confusão para duas classes.

		Classe de Referência	
		A	B
Classe Prevista	A	Verdadeiro Positivo (VP)	Falso Positivo (FP)
	B	Falso Negativo (FN)	Verdadeiro Negativo (VN)

No caso de um classificador perfeito, que acerta corretamente nas classes de todos os elementos, teríamos o total de elementos dividido entre as células VP e VN enquanto que em FN e FP o valor seria zero. No entanto na grande maioria das situações os FN e FP são diferentes de zero, e dependendo do problema em questão o foco poderá ser tentar reduzir o número de FP ou o número de FN.

### 3.5.2 Precisão

A precisão é uma métrica que apresenta a proporção de elementos que foram corretamente classificados como sendo de uma determinada classe, em relação ao total de elementos que foram classificados como sendo dessa classe.

$$Precisão = \frac{VP}{VP + FP}$$

### 3.5.3 Recall

O Recall é semelhante à precisão e apresenta a proporção de elementos que foram corretamente classificados como sendo de uma determinada classe, em relação ao total de elementos de referência que são efetivamente dessa classe.

$$Recall = \frac{VP}{VP + FN}$$

### 3.5.4 F1-score

O F1-score é uma métrica calculada a partir da precisão e do *recall*. Esta métrica pode ser vista como uma média ponderada entre a precisão e o *recall*, que sintetiza os valores destas métricas num só valor.

$$F1 = 2 * \frac{precisão * recall}{precisão + recall}$$

### 3.5.5 Accuracy

A accuracy é uma medida simples que é bastante usada e é calculada usando os elementos corretamente classificadas sobre o total de elementos. No entanto é necessário ter em conta a distribuição das classes ao analisar os valores desta métrica.

$$Accuracy = \frac{VP + VN}{VP + VN + FP + FN}$$

### 3.5.6 Kappa

O *Kappa* é uma métrica que mede a concordância entre as classificações geradas e as classes de referência e que tem em conta o facto desta concordância (ou discordância) ter sido por acaso.

$$kappa = \frac{p_o - p_e}{1 - p_e}$$

Nesta fórmula  $p_o$  representa a accuracy e  $p_e$  a probabilidade de a concordância ter sido por acaso.  $p_e$  é calculado através da seguinte fórmula:

$$p_e = \frac{1}{N^2} \sum_k n_{k1} * n_{k2}$$

sendo  $k$  as classes existentes,  $n_{k1}$  quantos elementos são efetivamente da classe  $k$  e  $n_{k2}$  quantos elementos foram classificados como sendo da classe  $k$ .

### 3.6 Conclusões

Nesta secção foi abordado o tema da aprendizagem automática aplicado ao caso concreto da deteção remota.

Foram descritos alguns algoritmos mencionados na literatura e que têm mostrado ao longo do tempo obter bons resultados em aplicações relacionadas com a classificação de vegetação ou da cobertura do terreno. De destacar o algoritmo RF que é usado em vários trabalhos por já ter mostrado ser robusto na resolução deste tipo de problemas.

Nos trabalhos desta área podem ser usadas duas abordagens uma que analisa os dados das imagens píxel a píxel e outra que tem por base a criação de objetos. Geralmente a abordagem baseada em objetos obtém melhores resultados que a abordagem baseada em píxeis, além de ter a vantagem dos tempos de computação serem inferiores aos da abordagem por píxeis.

No que diz respeito à validação dos resultados dos classificadores, os dados disponibilizados são extremamente importantes para avaliar a performance dos algoritmos, para tal existem várias métricas de validação, sendo que uma forma muito clara que permite avaliar o desempenho dos classificadores por classe é através de uma matriz de confusão.

Este conjunto de abordagens e algoritmos fornece uma boa base de informação sobre os métodos usados em deteção remota, que permite delinear uma abordagem indicada para o problema da deteção de intervenções nas FGCI.



## ABORDAGEM

## 4.1 Introdução

Neste capítulo é apresentada a metodologia seguida neste trabalho para processar os diferentes tipos de dados e são descritas as fases (Figura 4.1) que fazem parte deste processamento.

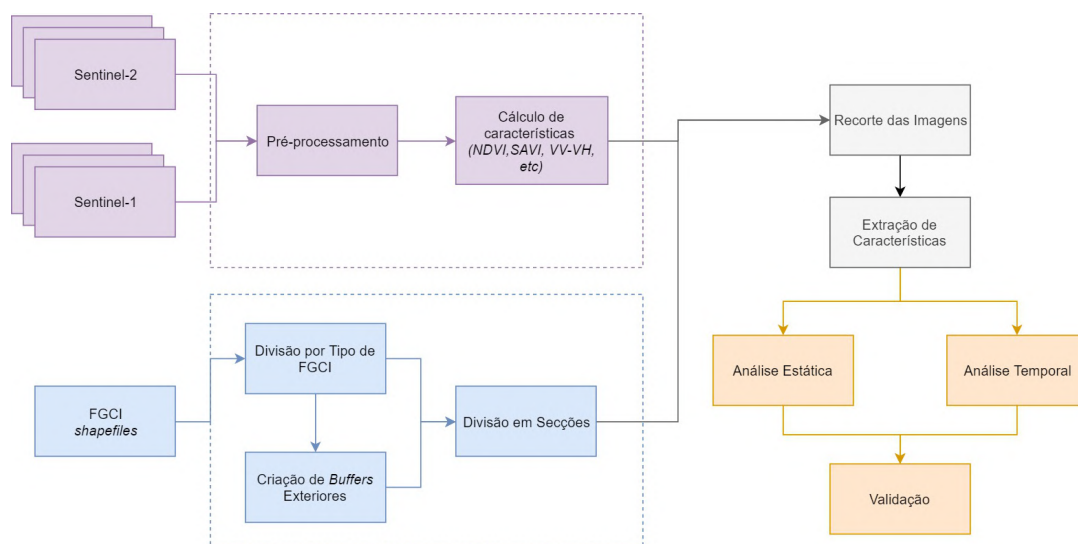


Figura 4.1: Diagrama que representa os passos da abordagem seguida.

As imagens satélite e a informação vetorial necessitam de ser pré-processadas de forma a que a informação a ser analisada seja o mais correta possível. Para tal é realizado um pré-processamento destas imagens de forma a eliminar possíveis erros de causados por fatores atmosféricos ou outros.

Através da informação vetorial das FGCI são determinadas áreas de interesse, que

dizem respeito às proximidades das FGCI e à própria faixa, que depois são usadas para extrair dados das imagens satélite já pré-processadas.

Com isto, os dados a serem utilizados são agrupados em diferentes conjuntos de forma a comparar o desempenho das diferentes características recolhidas. Assim, é descrito o conjunto de dados e características que foram usados e qual foi a abordagem utilizada para realizar a análise estática e análise temporal para estes conjuntos de dados.

Tendo por base o diagrama da abordagem na Figura 4.1, no restante deste capítulo é feita uma descrição geral do intuito de cada uma das fases da abordagem, ilustrando com alguns exemplos.

### 4.2 Pré-processamento de Imagens Satélite

As imagens provenientes de satélites têm propriedades diferentes que variam de satélite para satélite, ou até mesmo de órbita para órbita. Desta forma é necessário que exista um cuidado no processamento destas imagens.

Vários tipos de anomalias podem ocorrer quando uma imagem está a ser captada, desde falhas instrumentais a erros de processamento ou interferências atmosféricas. Um exemplo comum é a ocorrência de nuvens que podem cobrir grande parte da imagem captada.

A composição da atmosfera também pode afetar ligeiramente o sinal recebido, fazendo com imagens tiradas em momentos ou lugares diferentes sejam afetadas de diferentes formas. Outra situação que ocorre é que duas imagens da mesma zona, mas captadas em momentos diferentes, podem conter erros de georeferenciação, fazendo com um mesmo ponto esteja em localizações diferentes nas duas imagens.

Estas situações podem resultar em erros de classificação ou em modelos imprecisos, assim para minimizar possíveis erros é fundamental realizar um pré-processamento das imagens.

#### 4.2.1 Sentinel-1

Nas imagens do Sentinel-1 são realizadas algumas correções de forma a minimizar erros que possam ocorrer, utilizando ferramenta SNAP que é disponibilizada pela ESA.

A primeira correção realizada é a aplicação do ficheiro de órbita do satélite. Este ficheiro contém informação precisa da órbita que o satélite realizou quando captou a imagem, melhorando assim a geocodificação da imagem.

De seguida é realizada uma calibração que converte os valores digitais dos píxeis para valores de retrodispersão.

Devido às variações topográficas do terreno as distâncias representadas nas imagens captadas por satélites do tipo SAR podem sofrer distorções. Para minimizar estas distorções é realizada uma correção do terreno usando dados de elevação do terreno.



### 4.2.2 Sentinel-2

As imagens do Sentinel-2 por vezes têm imprecisões de georreferenciação fazendo com que uma localização se encontre incorretamente referenciada. Estas falhas, em alguns casos, podem chegar a algumas dezenas de metros [47].

Uma vez que algumas FGCI têm uma dimensão de 10m, que coincide com a resolução do Sentinel-2, por vezes ocorrem imprecisões na georreferenciação que se não forem corrigidas podem ter impacto nos resultados, fazendo com que estes sejam referentes a áreas que não dizem respeito ao interior da faixa.

Assim, é necessário alinhar as imagens para minimizar este tipo de erros. Para tal, foi criada uma imagem base, composta por vários produtos do Sentinel-2, que cobre todo território português. Cada produto foi alinhado manualmente, tendo como referência visual o OpenStreetMap e o Google Maps.

Tendo este conjunto de imagens como referência, todas as imagens descarregadas são alinhadas em relação à imagem de referência. Para realizar o alinhamento foi usada a biblioteca OpenCV [28] que disponibiliza ferramentas que permitem identificar pontos em comum nas duas imagens e realizar transformações em imagens.

### 4.2.3 Reamostragem de Imagens

Algumas bandas do satélite Sentinel-2 têm uma resolução máxima de 20m ou 60m em bandas do infravermelho próximo e no infravermelho de onda curta.

Alguns dos índices de vegetação usados utilizam bandas que têm resolução de 10m e de 20m. Assim, para realizar os cálculos que permitem obter os valores destes índices foi realizada uma reamostragem das bandas de 20m, passando estas para 10m.

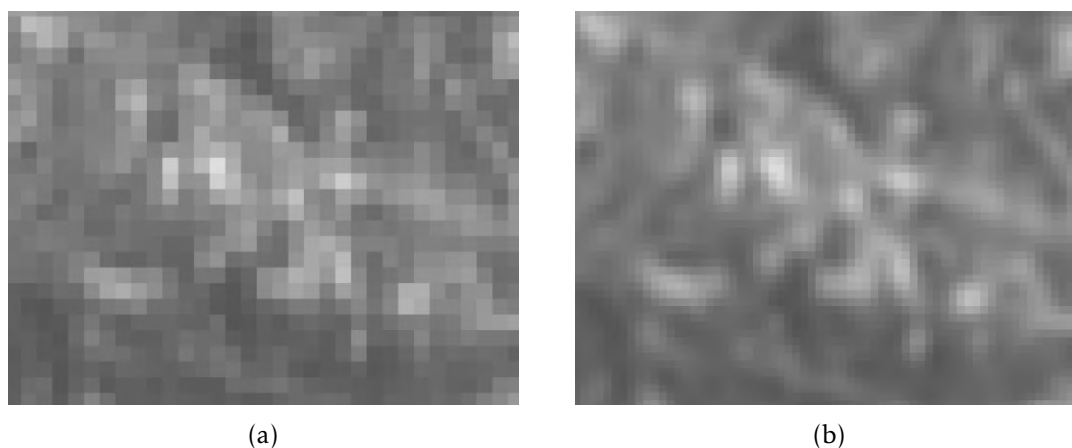


Figura 4.2: Reamostragem da banda B07 do Sentinel-2 de 20m para 10m. (a) Imagem com 20m de resolução. (b) Imagem reamostrada, com 10m de resolução.

Esta reamostragem foi realizada usando o GDAL, mais concretamente a ferramenta *warp*. Foram visualmente comparados diferentes métodos de amostragem (*near*, *bilinear*,

*cubic*, *cubicspline*, *lanczos*) usando imagens que têm resolução espacial de 10m e que também são disponibilizadas com resoluções de 20m e 60m. Foi realizada a reamostragem, usando os vários métodos, numa imagem de 20m e os resultados foram comparados com a imagem de 10m.

O método que mostrou resultados mais semelhantes à resolução de 10m foi o *lanczos*, sendo que este foi o método escolhido para realizar a reamostragem das bandas de 20m.

### 4.3 Determinação das Zonas de Interesse das FGCI

Os dados vetoriais das FGCI disponibilizados pelo ICNF, são usados para extrair das imagens satélite informação relativa às faixas.

Além da informação referente às faixas e o seu interior, foi usada informação da zona exterior e adjacente às faixas (*buffers*) para permitir comparar a vegetação nas duas áreas. A criação destes *buffers* foi efetuada recorrendo a bibliotecas que permitem realizar vários tipos de operações sobre dados vetoriais e também dados raster.

Nesta secção é descrito o processamento realizado à informação vetorial, para que possa ser extraída informação de interesse das imagens satélite e analisados os diferentes tipos de FGCI.

#### 4.3.1 FGCI

Para analisar individualmente faixas de diferentes tipos, foi necessário separar as faixas pelo o seu tipo. Desta forma, houve uma divisão prévia em 4 tipos diferentes: faixas em redor de habitações e outras infraestruturas; faixas ao longo de estradas; faixas pertencentes à rede primária de FGCI; e faixas ao longo de linhas de alta e média tensão.

Esta separação permite analisar cada tipo de faixa de forma independente e ainda facilitar outros processos, como por exemplo o de criação de *buffers* exteriores, que será descrito posteriormente.

#### 4.3.2 Buffers

A abordagem proposta para detetar se uma determinada faixa foi intervencionada passa por comparar a vegetação no interior da faixa com a vegetação no exterior adjacente à faixa.

Para tal, foram criados *buffers* exteriores às FGCI com 20m metros de largura. Com estes *buffers*, e os polígonos das faixas, pode ser extraída informação das imagens satélite e comparar dados do exterior com dados do interior da FGCI. Ao comparar dados do exterior próximo à faixa, com dados do interior da faixa, se existir algum tipo de intervenção na faixa, este pode ser detetado.

Na Figura 4.3 podem ser observados exemplos de *buffers* para os diferentes tipos de faixas.

Como a forma dos polígonos que representam as FGCI varia de acordo com o tipo de faixa, o processo de criação destes *buffers* não é igual para todas as faixas.

Faixas que têm zonas no seu interior que não fazem parte da faixa em si, como é caso das estradas (Figura 4.3c) e das habitações (Figura 4.3a), quando é criado um *buffer* este também cobre as zonas interiores. É necessário remover os *buffers* criados no interior uma vez que são zonas onde não existe vegetação para comparar.

Nas faixas ao longo de linhas elétricas (4.3d) e nas faixas da RPFGC (4.3b) a criação do *buffer* exterior é mais simples pois não tem zonas interiores que não fazem parte da faixa.



(a) *Buffer* exterior nas FGCI ao redor de habitações.



(b) *Buffer* exterior nas faixas da rede primária.



(c) *Buffer* exterior nas FGCI ao longo de estradas.



(d) FGCI linhas elétricas e o *buffer* exterior.

Figura 4.3: Exemplos de *buffers* para os diferentes tipos de faixas.

Na criação dos *buffers* foi tido em conta as zonas em que estes intersectam com estradas ou com outra FGCI, e como tal as zonas de intersecção foram removidas de forma a tentar garantir que os *buffers* apenas dizem respeito a zonas de vegetação. Em todos os *buffers* da Figura 4.3 podem ser vistos exemplos onde ocorrem interseções com estradas e onde a zona de intersecção foi removida.

Na secção 5.3 encontra-se detalhado o processo de criação dos *buffers* exteriores para

os diferentes tipos de faixa.

### 4.3.3 Divisão em Secções

Para analisar as FGCI por segmentos foi usada uma abordagem por objetos dividindo as FGCI em secções com aproximadamente a mesma dimensão.

Uma vez que a resolução espacial das imagens é de 10m, algumas faixas podem ter vários píxeis de largura, fazendo com que seja necessário comparar os valores dos vários píxeis do interior da faixa com os valores dos píxeis do *buffer* exterior, o que pode ser complexo de implementar. A abordagem recorrendo ao uso de objetos facilita bastante este processo de comparação, fazendo com que só existam dois valores para comparar, o de dentro da faixa e do *buffer* exterior.

O uso desta abordagem também permite reduzir o ruído presente nos píxeis do interior das faixas, que podem estar contaminados com valores de refletância de estruturas adjacentes à faixa (e.g. asfalto, no caso das FGCI ao longo de estradas). Ao usar objetos, cada objeto é composto por um conjunto de píxeis (uns com menos ruído que outros) sobre o qual é calculada a média dos seus valores, fazendo com que píxeis mais afetados pelo ruído não influenciem tanto os resultados. Outra vantagem desta abordagem é a redução do conjunto de dados, diminuindo os tempos de execução de algoritmos de aprendizagem automática e outros.

A divisão em secções é aplicada em todos os tipos de faixas após a criação dos *buffers*, assim cada secção tem informação relativa ao interior da FGCI e ao *buffer* exterior (Figura 4.4).

O método usado para realizar esta divisão passa por transformar os píxeis que correspondem ao interior e *buffer* exterior da faixa em pontos (Figura 5.9a). De seguida é corrido o algoritmo *K-Means* com estes pontos.



(a) FGCI ao longo de estradas e o respetivo *buffer*. (b) Divisão da FGCI e do *buffer* em secções.

Figura 4.4: Divisão das FGCI ao longo das estradas e dos seus *buffers* em secções

Foram analisadas principalmente faixas ao longo de estradas e linhas elétricas pois estas foram as faixas que foram intervenções pelo município de Mação e para as quais foi disponibilizada informação. Também foram realizadas algumas experiências nas faixas em redor às habitações, embora não existissem dados sobre as intervenções. Como tal, a divisão em secções só foi feita para estes três tipos de faixas.

Para ser possível extrair dados de apenas uma secção e analisar o comportamento da vegetação no *buffer* e no interior da FGCI dessa secção, cada secção é intersectada com a FGCI e com o *buffer* e o resultado da intersecção é guardado. Na Figura 4.5 está um exemplo da intersecção de uma secção com a FGCI e *buffer* para diferentes tipos de faixa.

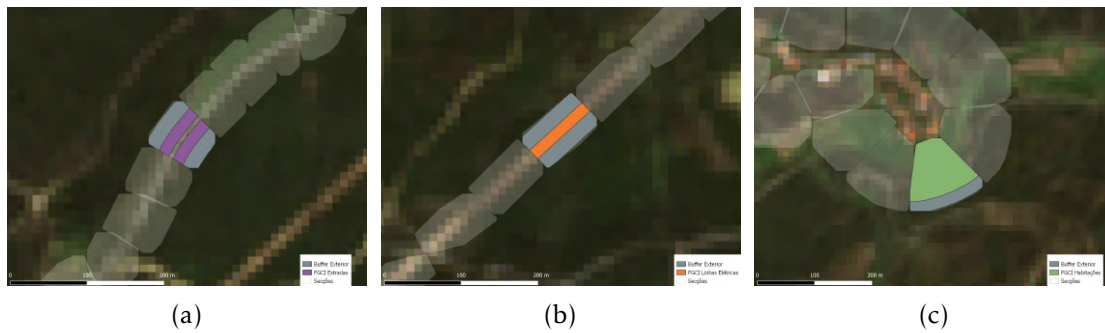


Figura 4.5: Exemplos da intersecção das FGCI e dos *buffers* com as secções criadas. (a) Secção de uma FGCI ao longo de uma estrada. (b) Secção de uma FGCI ao longo de uma linha elétrica. (c) Secção de uma FGCI ao redor de um aglomerado de habitações.

## 4.4 Conjunto de Dados

Diferentes características foram usadas para descrever o interior e *buffer* exterior de cada uma das secções.

Foram usados dados dos satélites Sentinel-1 e Sentinel-2 obtidos ao longo do ano de 2018. Além dos dados das bandas e das polarizações foram extraídas novas características, na Tabela 4.1 lista-se a informação adicional que foi calculada para cada um dos satélites. Para o Sentinel-1 foi calculada a diferença e o rácio entre a polarização VV e a VH, uma vez que já mostrou melhorar a exatidão das classificações [1]. Para o Sentinel-2 foram usados os índices de vegetação descritos na secção 2.4.

Tabela 4.1: Características calculadas a partir dos dados de cada um dos satélites.

Satélite	Características Calculadas
Sentinel-1	VV / VH, VV - VH
Sentinel-2	NDVI, NDWI, SAVI, IRECI, SR

Tendo por base todas estas características foram criados 4 conjuntos de dados de forma a comparar o impacto que as diferentes características têm no resultado (Tabela



4.2). O conjunto de dados *CD\_TUDO* contém todas as bandas do Sentinel-2, índices de vegetação e polarizações, a partir deste conjunto foram criados outros dois o *CD\_BANDAS*, que apenas tem as bandas do Sentinel-2 e o *CD\_ÍNDICES* que apenas tem índices de vegetação de forma a comparar a performance das bandas com a dos índices. Finalmente foi criado o conjunto de dados *CD\_FUSÃO* que contém as bandas no verde, vermelho, *red-edge* e infravermelho próximo que são usadas nos índices de vegetação e dois índices de vegetação o NDVI e NDWI por estimarem características diferentes da vegetação (o vigor e o conteúdo de água).

Para cada uma das características é extraída informação do *buffer* exterior e do interior da FGCI. Além da informação do *buffer* e da faixa, é ainda calculada a diferença (equação 4.1) e o rácio (equação 4.2) entre o *buffer* exterior e o interior da FGCI para cada características. Assim, para cada secção temos dados do *buffer* exterior, dados do interior das FGCI e a diferença e o rácio entre os dois.

$$dif_x = x_b - x_i \quad (4.1)$$

$$ratio_x = \frac{x_b}{x_i} \quad (4.2)$$

Onde  $x_b$  representa os dados do *buffer* da característica  $x$ , e  $x_i$  representa os dados do interior da FGCI da característica  $x$ .

Tabela 4.2: Conjuntos de dados usados e respetivas características.

Conjunto de Dados	Características Base
<i>CD_TUDO</i>	NDVI, NDWI, SAVI, IRECI, SR, B02, B03, B04, B05, B06, B07, B08, B11, (VV, VH, VV/VH, VV-VH)*
<i>CD_ÍNDICES</i>	NDVI, NDWI, SAVI, IRECI, SR, (VV, VH, VV/VH, VV-VH)*
<i>CD_BANDAS</i>	B02, B03, B04, B05, B06, B07, B08, B11, (VV, VH, VV/VH, VV-VH)*
<i>CD_FUSÃO</i>	NDVI, NDWI, B03, B04, B05, B06, B07, B08, (VV, VH, VV/VH, VV-VH)*

\* Apenas usado na análise temporal.

Apesar dos dados provenientes do satélite Sentinel-1 fazerem parte dos conjuntos de dados, estes apenas foram usados na análise temporal (i.e. quando foram feitas classificações recorrendo a séries temporais), e ainda assim nesta análise foram realizados testes com e sem estes dados.

## 4.5 Análise Estática

A análise estática consiste na avaliação de informação satélite que respeitante a apenas uma data. Para tal foram usadas imagens de uma data após terem sido realizadas as intervenções nas FGCI.

Nesta análise apenas foram usados dados do satélite Sentinel-2 uma vez que não foram obtidas imagens do Sentinel-1 que coincidissem com a mesma data que as do Sentinel-2.

Cada tipo de faixa foi analisado individualmente e foram usados os algoritmos de classificação mencionados na secção 3.2 e adicionalmente o algoritmo de agrupamento *K-Means*. Numa primeira fase, antes de ter sido obtida informação oficial sobre as faixas intervencionadas, foram realizadas experiências nas faixas ao longo de estradas usando o *K-Means* e o índice de vegetação NDVI.

Com a informação sobre as faixas que foram intervencionadas já disponível foram então usados algoritmos de classificação usando os conjuntos de dados referidos anteriormente, e analisadas as faixas ao longo de estradas e linhas elétricas.

Os resultados da análise estática além de serem indicadores da qualidade dos modelos e da capacidade destes se adaptarem ao problema em questão, também serviram de base de comparação para a análise temporal que foi realizada posteriormente.

## 4.6 Análise Temporal

Para realizar a análise temporal dos dados foram utilizadas as séries temporais das várias características e extraídas métricas, que posteriormente foram introduzidas nos algoritmos de aprendizagem automática descritos na secção 3.2.

Além de ser usada informação do interior da faixa, do *buffer*, da diferença e do rácio, foi calculado um novo indicador, o  $\Delta dif$  (equação 4.3). Este indicador usa a  $dif_x$  de duas datas consecutivas e faz a diferença entre ambas, quantificando o aumento ou declínio de uma dada característica entre as duas datas.

$$\Delta dif_k = dif_{x,k} - dif_{x,k-1} \quad , \quad k > 0 \quad (4.3)$$

Na expressão 4.3,  $dif_{x,k}$  representa a diferença entre o *buffer* e o interior da FGCI, para a característica  $x$ , na data  $k$ .

Com isto, temos padrões temporais de todas as características dentro da faixa, no *buffer* exterior, a diferença, o rácio e  $\Delta dif$ . Destes padrões são extraídas as métricas: mínimo, máximo, média e desvio padrão para cada uma das secções.

Esta informação inserida nos algoritmos de aprendizagem automática permite estabelecer se uma determinada secção de uma faixa foi ou não intervencionada.

## 4.7 Conclusões

Neste capítulo foi apresentada a abordagem definida para processar as imagens satélite e a informação vetorial das FGCI. Foram apresentadas as fases da abordagem e o objetivo de cada uma delas.

O pré-processamento das imagens é uma das primeira fases, que tem como finalidade corrigir eventuais erros presentes nas imagens satélite e prepará-las para depois poderem ser usadas.

A fase de determinação das zonas de interesse é fundamental para analisar as FGCI, é nesta fase onde são criados os *buffers* exteriores e onde é realizada a divisão das faixas em secções. Este processamento feito à informação vetorial é central para depois ser possível determinar se uma secção de uma faixa foi ou não intervencionada.

Finalmente foram identificados os conjuntos de dados usados para treinar os algoritmos de classificação e como é que estes foram criados, e apresentadas as duas análises feitas aos conjuntos de dados: a estática e a temporal.





## IMPLEMENTAÇÃO

### 5.1 Introdução

Neste capítulo é descrito detalhadamente cada uma das fases da implementação da abordagem usada para analisar as FGCI.

É explicada a fase do pré-processamento de imagem que realiza o alinhamento das imagens satélite, e que necessitou de ser implementada recorrendo a diferentes bibliotecas. A maioria das implementações apresentadas neste capítulo são referentes ao processamento da informação vetorial das FGCI, e têm como objetivo final obter as áreas que irão ser analisadas.

Todos os processos de pré-processamento das imagens satélite e do tratamento da informação vetorial das FGCI foram automatizados usando a linguagem de programação *Python* versão 3.7.1. Para o tratamento da informação vetorial, a principal biblioteca usada foi o GDAL/OGR versão 2.4.1 [16], sendo que o grosso da automatização foi em funções relacionadas com a manipulação de polígonos.

### 5.2 Alinhamento de Imagens

Todas as imagens, antes de serem analisadas, são alinhadas de forma a minimizar eventuais erros de georreferenciação que possam existir.

Para realizar o alinhamento foi usada a biblioteca *OpenCV* que disponibiliza ferramentas que permitem efetuar diferentes tipos de operações de manipulação de imagem.

O alinhamento é feito usando o algoritmo ORB<sup>1</sup> para detetar pontos homólogos que se destacam nas duas imagens (na imagem de referência e na imagem a ser alinhada), de

---

<sup>1</sup>[https://docs.opencv.org/3.0-beta/doc/py\\_tutorials/py\\_feature2d/py\\_orb/py\\_orb.html](https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_orb/py_orb.html)

seguida é usado o *DescriptorMatcher*<sup>2</sup> para criar uma correspondência entre esses pontos. Na Figura 5.1 pode ser visto um exemplo da correspondência de pontos homólogos entre duas imagens semelhantes.

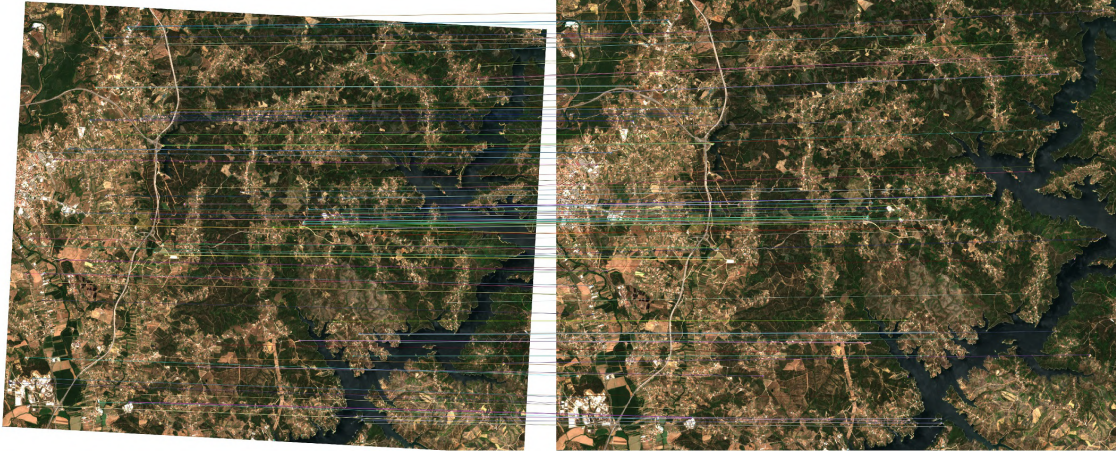


Figura 5.1: Correspondência de pontos entre duas imagens semelhantes, sendo que a imagem da esquerda apenas tem uma rotação.

A cada par de pontos homólogos está associada a distância entre eles. Os pares de pontos são ordenados por ordem ascendente usando a distância entre pontos, e é selecionada uma percentagem de pontos que apresenta a menor distância (neste caso foram usados 25%).

Com os pontos homólogos selecionados é calculada a transformação afim, recorrendo à função *estimateAffine2D*<sup>3</sup>. Esta função calcula uma matriz (matriz 5.1), que ao aplicar na imagem a ser alinhada, realiza uma translação, rotação ou ampliação na imagem.

$$M = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \quad (5.1)$$

Desta matriz são extraídos apenas os valores que dizem respeito às translações, representados na matriz por  $x_{13}$  e  $x_{23}$ , onde  $x_{13}$  representa a translação em  $x$  e  $x_{23}$  em  $y$ . Estes valores são representados em píxeis, sendo que depois são transformados em valores do deslocamento a ser aplicado às coordenadas geográficas da imagem desalinhada, usando as expressões 5.2 e 5.3.

$$dx = -x_{13} \times 10 \quad (5.2)$$

$$dy = x_{23} \times 10 \quad (5.3)$$

<sup>2</sup>[https://docs.opencv.org/3.4/db/d39/classcv\\_1\\_1DescriptorMatcher.html](https://docs.opencv.org/3.4/db/d39/classcv_1_1DescriptorMatcher.html)

<sup>3</sup>[https://docs.opencv.org/3.4/d9/d0c/group\\_\\_calib3d.html#ga27865b1d26bac9ce91e\faee83e94d4dd](https://docs.opencv.org/3.4/d9/d0c/group__calib3d.html#ga27865b1d26bac9ce91e\faee83e94d4dd)

Nestas expressões a constante 10 representa a resolução espacial das imagens satélite.

A alteração das coordenadas geográficas da imagem desalinhada é feita usando a biblioteca GDAL, usando a função *GetGeoTransform* para obter as coordenadas antigas e alterá-las usando os valores  $dx$  e  $dy$ . Por fim, os novos valores das coordenadas são atualizados usando a função *SetGeoTransform*.

### 5.3 Informação vetorial

O processamento dos dados vetoriais das FGCI envolve vários passos. Estes passos visam facilitar o processamento e análise dos diferentes tipos de FGCI, sendo que o passo fundamental e que é a motivação deste processamento é o de criação de *buffers*.

A Figura 5.2 apresenta todos os passos que fazem parte do processamento da informação vetorial. Nesta secção é descrita a implementação de cada um destes passos, sendo que alguns deles são ilustrados com excertos do código usado.

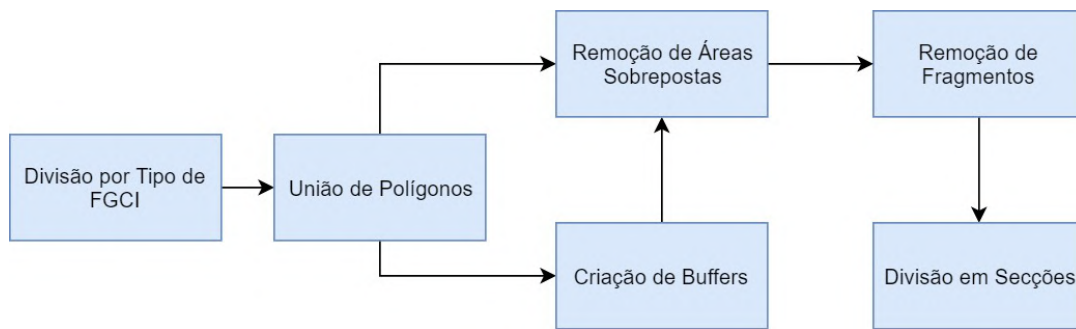


Figura 5.2: Cadeia de passos para o processamento das FGCI e criação dos *buffers*.

#### 5.3.1 Divisão das FGCI por Tipo e União de Polígonos

Os 4 tipos de faixas nos quais o ficheiro original foi dividido são: ao redor de habitações e aglomerados habitacionais, ao longo de linhas elétricas, ao longo de estradas e faixas pertencentes à RPF GCI. Estas faixas foram separadas em 4 ficheiros diferentes, de forma a facilitar a análise de cada uma delas.

O processo de separação envolve a criação dos novos ficheiros, a identificação do tipo de faixa através do atributo *DESC\_FGC* e a cópia do polígono para o novo ficheiro.

Com os diferentes tipos faixas divididos em polígonos segue-se o processo de união de geometrias. Ao realizar operações de interseção ou diferença entre polígonos, por vezes, nas linhas de separação entre polígonos (na Figura 5.3a podem ser vistas algumas destas linhas), são criados alguns artefactos indesejados. Para evitar estes artefactos, é realizada uma união prévia de todos os polígonos.

Pode ser visto no exemplo 5.1 o código usado para realizar a união dos polígonos. Este código recorre à função do *VectorTranslate*, do GDAL, que usa funcionalidades da

biblioteca *Spatialite* (extensão do sistema de base de dados *SQLite*, que disponibiliza funcionalidades espaciais) para realizar *queries* aos elementos de um determinado *shapefile*. Na Figura 5.3b pode ser observado o resultado da união dos polígonos nas faixas ao longo de estradas.

Listagem 5.1: Código usado para unir os Polígonos.

```

1 def merge_geometries_by_field(file,out, field):
2
3     file_name = os.path.basename(file)
4
5     if field == None or field == '':
6         query = "SELECT ST_Union(geometry) AS geometry FROM '" + file_name
7             [:-4]+'"'
8     else:
9         query = "SELECT ST_Union(geometry) AS geometry, '"+field+"' FROM '" +
10             file_name[:-4]+'"' GROUP BY '"+field"
11
12     ds = gdal.VectorTranslate(out,file,
13                             SQLDialect="sqlite",
14                             SQLStatement=query,
15                             format="ESRI Shapefile",
16                             geometryType="PROMOTE_TO_MULTI")
17
18     ds = None

```

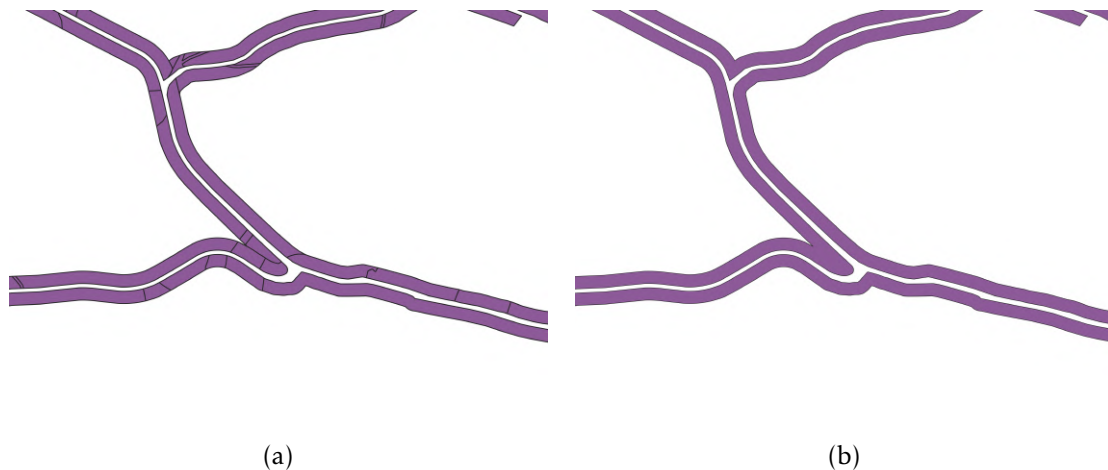


Figura 5.3: Exemplo do resultado da união de polígonos (faixas ao longo de estradas). (a) Formato original da FGCI. (b) FGCI após união de polígonos.

### 5.3.2 Criação de Buffers

Nesta secção é descrito o processo de criação dos *buffers* exteriores para os diferentes tipos de faixas.

### 5.3.2.1 FGCI Linhas Elétricas, RPFGC e Estradas

O processo de criação de *buffers* mais simples é o das faixas ao longo de linhas elétricas, uma vez que os polígonos que representam estas faixas têm um formato retangular (Figura 5.4a). Assim, para cada polígono deste tipo de faixa é executado o método *Buffer*, da biblioteca GDAL, resultando um polígono, com a mesma forma, apenas com mais 20m para cada um dos lados (Figura 5.4b).

De seguida é realizada a diferença entre o polígono original da faixa (Figura 5.4a) e o novo (Figura 5.4b) obtemos o *buffer* exterior de 20m. A função criada para realizar a operação de diferença pode ser vista na Listagem 5.2, esta função recebe dois ficheiros, um com a FGCI sem *buffer* e outro da FGCI com *buffer*. É realizada uma união prévia dos polígonos de ambos os ficheiros e de seguida é que é realizada a diferença entre esses polígonos.

Listagem 5.2: Código usado realizar a diferença entre polígonos.

```

1 def difference(f1, f2, out):
2     try:
3         fgc_shp = ogr.Open(f1)
4         lyr1 = fgc_shp.GetLayer()
5
6         fgc_shp2 = ogr.Open(f2)
7         lyr2 = fgc_shp2.GetLayer()
8
9         driver=ogr.GetDriverByName('ESRI Shapefile')
10        ds=driver.CreateDataSource(out)
11        diff_lyr = ds.CreateLayer('temp', lyr2.GetSpatialRef(), ogr.
            wkbMultiPolygon )
12
13        #uniao de todos os poligonos em ambos os ficheiros
14        union1 = ogr.Geometry(ogr.wkbMultiPolygon)
15        for feat1 in lyr1:
16            geom1 = feat1.GetGeometryRef()
17            if geom1 != None:
18                union1 = union1.Union(geom1)
19
20        union2 = ogr.Geometry(ogr.wkbMultiPolygon)
21        for feat2 in lyr2:
22            geom2 = feat2.GetGeometryRef()
23            if geom2 != None:
24                union2 = union2.Union(geom2)
25
26        #buffer com dimensao zero, para corrigir eventuais erros nas
27        geometrias
28        union1 = union1.Buffer(0)
29        union2 = union2.Buffer(0)
30
31        #Difrenca entre os poligonos unidos dos dois ficheiros
32        diff = union1.Difference(union2)

```

```

32
33     #Guardar o resultado num novo ficheiro
34     new_feat = ogr.Feature(diff_lyr.GetLayerDefn())
35     new_feat.SetGeometry(diff)
36     diff_lyr.CreateFeature(new_feat)
37 except:
38     print("exception thrown!")
39     traceback.print_exc()

```

Na Figura 5.4c pode ser visto um exemplo do resultado final da criação de *buffers* exteriores a faixas de linhas elétricas de alta e média tensão, já com as zonas de intersecção com estradas e outras faixas removidas.

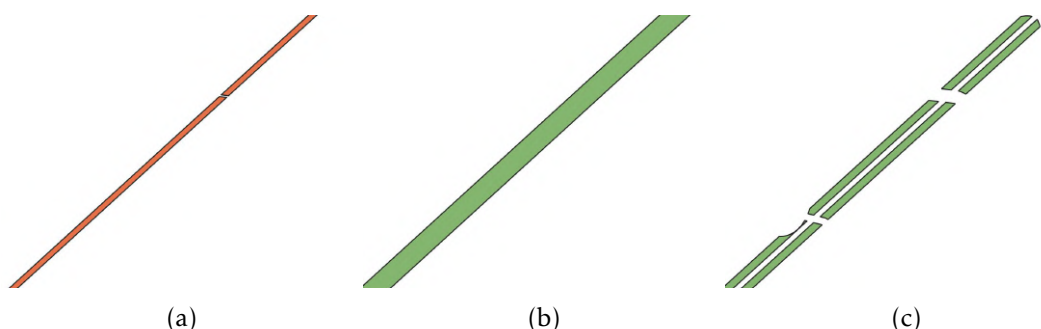


Figura 5.4: Fases da criação dos *buffers* exteriores. (a) Polígono que representa uma FGCI ao longo de linhas elétricas. (b) Polígono resultante do uso da função *Buffer* com 20m. (c) Polígonos que representam apenas áreas exteriores à FGCI.

Apesar dos polígonos das FGCI pertencentes à rede primária não terem a mesma forma geométrica, nem as mesmas dimensões, que os das linhas elétricas, a criação de *buffers* exteriores para estes tipos de faixa é igual. No caso das FGCI ao longo das estradas o processo também é semelhante, só tem um passo extra que consiste na remoção das zonas do *buffer* que sobrepõem a estrada que passa no meio das faixas, e que irá ser abordado posteriormente na secção 5.3.3.

### 5.3.2.2 FGCI Habitações

Nas faixas em redor das habitações e aglomerados habitacionais a criação destes *buffers* envolve mais passos, isto porque os polígonos destas faixas têm uma forma circular, em que no centro estão as habitações (Figura 5.6a). Ao usar a função *Buffer* neste tipo de faixas, é gerado um polígono que sobrepõe zonas interiores que não podem ser usadas para comparação de vegetação (i.e. casas e estradas), um exemplo do resultado da aplicação da função *Buffer* nestas faixas pode ser visto na Figura 5.5.

Para remover as zonas interiores, que sobrepõem casas, é necessário preencher toda a área interior que diz respeito a zonas de habitação e fazer a diferença entre essa área e o polígono gerado pela função *Buffer*.



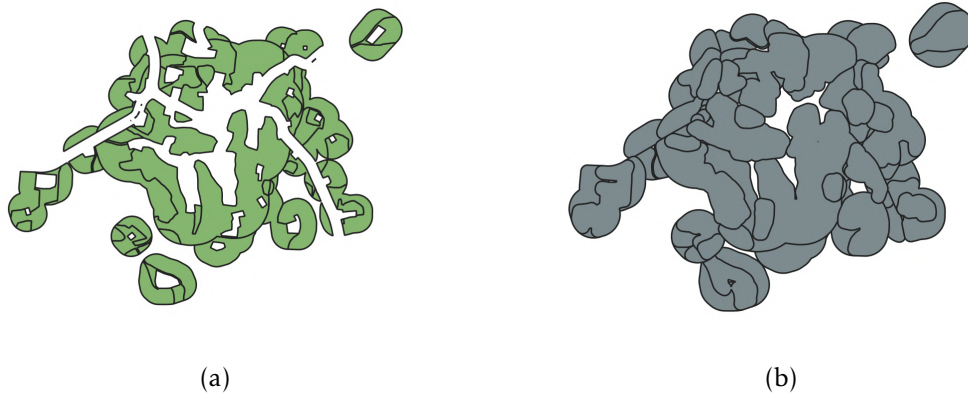


Figura 5.5: Exemplo do resultado da aplicação da função *Buffer* em algumas faixas ao redor de habitações. (a) Polígonos originais que representam a FGCI. (b) Polígonos resultantes após a criação de um *buffer* de 20m.

Para preencher o interior é necessário extrair todos os pontos que fazem parte dos polígonos da faixa, para tal é usado o identificador único da faixa *ID\_R\_FGC* de forma a identificar todos os polígonos associados a essa faixa e extrair os respectivos pontos. Na Listagem 5.3 pode ser visto o código usado para realizar a extração, que recebe um ficheiro com apenas as faixas em redor das habitações.

Listagem 5.3: Extração de pontos das faixas ao em redor das habitações.

```

1 def get_points_from_geometry(file):
2     fgc_shp = ogr.Open(file)
3     lyr = fgc_shp.GetLayer()
4
5     points_by_id = pd.DataFrame(data=[], columns=["id", "points"])
6
7     for feature in lyr:
8         i_desc_fgc = feature.GetFieldIndex("ID_R_FGC")
9         fgc_id = feature.GetFieldAsInteger(i_desc_fgc)
10        geom = feature.GetGeometryRef()
11
12        for j in range(0, geom.GetGeometryCount()):
13            ring = geom.GetGeometryRef(j)
14            if ring != None:
15                #Se a faixa ainda nao foi vista adiciona ao dataframe
16                if (points_by_id[points_by_id['id']==fgc_id].shape[0] == 0):
17                    points_by_id = points_by_id.append({'id' : fgc_id , '
18                                                         points' :[] } , ignore_index=True)
19
20                #Adicionar os pontos a lista de pontos da respectiva faixa
21                for i in range(0, ring.GetPointCount()):
22                    lon, lat, z = ring.GetPoint(i)

```



```

22         points_by_id[points_by_id['id']==fgc_id].iloc[0,1].append
23             ((lon,lat))
24     return points_by_id

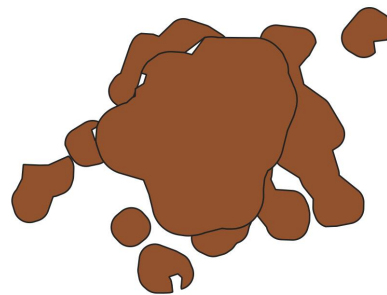
```

Com estes pontos é executada uma implementação<sup>4</sup> do algoritmo *Concave Hull* [25], resultando no conjunto de polígonos da Figura 5.6b.

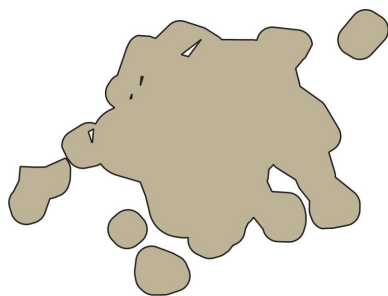
Ao executar o *Concave Hull* por vezes ficam a faltar partes que pertencem à FGCI e é necessário fazer uma união com o polígono original da FGCI, unido também os polígonos adjacentes num só (Figura 5.6c). De seguida são eliminados os buracos resultantes, ficando a FGCI com o interior preenchido (Figura 5.6d).



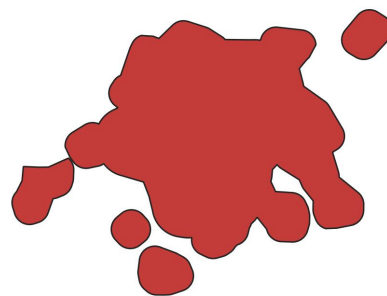
(a) FGCI inicial



(b) *Concave Hull* aplicado



(c) Geometrias unidas



(d) Buracos Eliminados

Figura 5.6: Processo para obter o interior das FGCI ao redor de habitações.

Finalmente é feita a diferença entre o polígono gerado pela função *Buffer* (Figura 5.5b) e a FGCI com o interior preenchido (Figura 5.6d), resultando apenas o *buffer* exterior à faixa (Figura 5.7).

<sup>4</sup><https://github.com/sebastianbeyer/concavehull>



Figura 5.7: Resultado da diferença entre a FGCI com um buffer de 20m (Figura 5.5b) e o polígono que representa o interior da FGCI (Figura 5.6d).

### 5.3.3 Remoção de Áreas Sobrepostas e de Fragmentos

Após a criação dos *buffers* algumas zonas sobrepõem áreas que pertencem ao interior de outras FGCI, e além disso também existem zonas em que estes *buffers* intersectam com estradas. Uma vez que é importante que os *buffers* apenas digam respeito a áreas de vegetação, é necessário remover estas zonas.

Apresenta-se na Figura 5.8a um exemplo de uma intersecção em que o *buffer* no exterior da FGCI ao longo de uma linha elétrica intersecta o interior da FGCI ao longo de uma estrada e outras estradas que não fazem parte da rede de faixas.

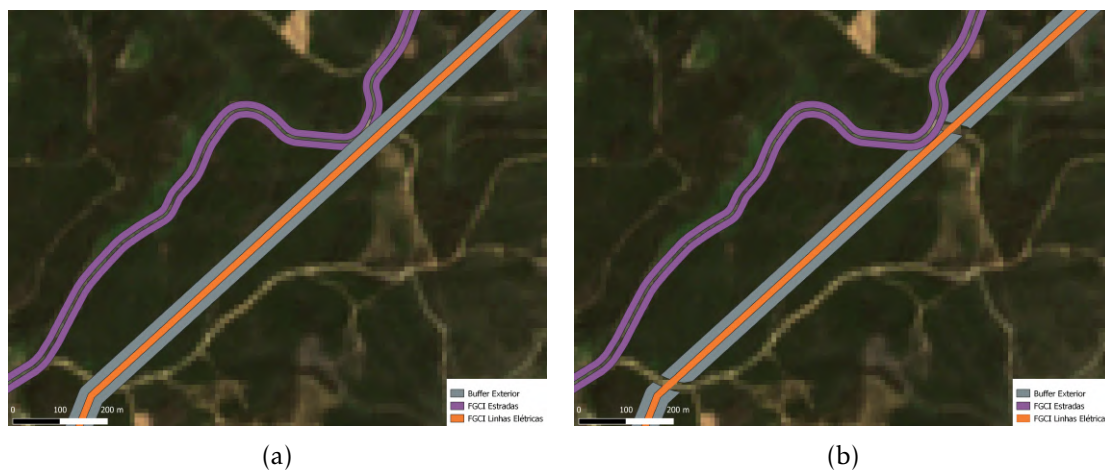


Figura 5.8: Exemplo da remoção de áreas de sobreposição com outras faixas. (a) Sobreposição do *buffer* com FGCI ao longo de estradas. (b) *buffer* resultante da remoção de zonas de sobreposição.

Para remover as zonas de intersecção com outras FGCI é realizada a diferença entre o *buffer* e todas as outras FGCI. No caso das estradas é usada a informação vetorial das

estradas de Portugal disponibilizada pelo OpenStreet Maps <sup>5</sup>, e é feita a diferença entre o *buffer* e a informação vetorial das estradas. O resultado desta remoção é visível na Figura 5.8b.

Ao realizar as operações de diferença entre polígonos por vezes resultam polígonos que têm uma dimensão reduzida que por vezes pode ser inferior a um píxel. Uma vez que estes polígonos de reduzida dimensão podem gerar alguns problemas no momento da extração da informação das imagens, todos os polígonos com um área inferior a  $200m^2$ , o que equivale a 2 píxeis de uma imagem do Sentinel-2, são removidos.

### 5.3.4 Divisão das FGCI em secções

O método usado para realizar esta divisão passa por transformar os píxeis que correspondem ao interior e *buffer* exterior da faixa em pontos (Figura 5.9a). De seguida é aplicado o algoritmo *K-Means* com estes pontos.

A finalidade do *K-Means* é criar aglomerados de pontos com base na distância entre pontos, de forma gerar *clusters* contíguos, com aproximadamente a mesma dimensão. O valor do parâmetro K é definido dividindo o número total de pontos pelo número de pontos que pretendemos que cada *cluster* tenha.

No caso das FGCI ao longo das estradas, temos um total de 208 mil pontos, e foram usados 40 pontos por *cluster*, resultando em 5190 *clusters* ou secções. O número de pontos por secção não pode ser muito baixo, pois pode gerar secções que não cobrem a faixa e o respetivo *buffer* exterior, nem pode ser muito alto para reduzir a possibilidade de uma secção conter zonas limpas e não limpas.

Finalmente, para obter o polígono que corresponde a uma secção, é corrido o algoritmo *Convex Hull* para cada um dos *clusters* gerados pelo *K-Means*. O resultado é um conjunto de polígonos com aproximadamente a mesma área, que cobrem a totalidade da FGCI (Figura 5.9c).

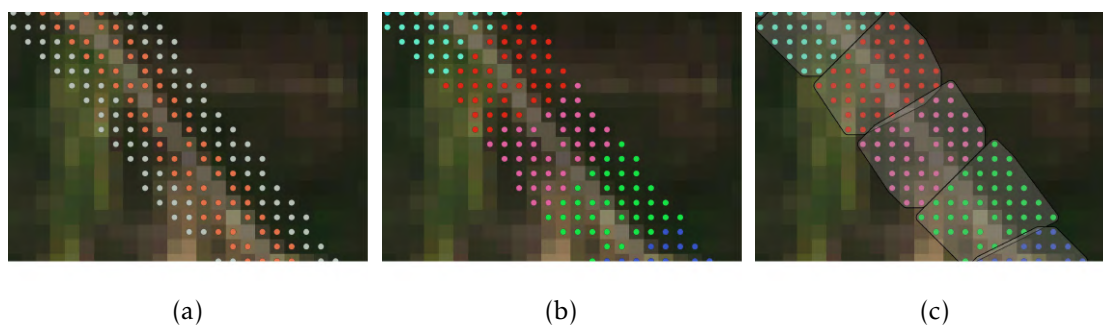


Figura 5.9: Fases da criação de secções. (a) Pontos do interior (laranja) e do *buffer* exterior (cinzento) da FGCI. (b) *Clusters* gerados usando K-Means. (c) Polígonos gerados para cada um dos *clusters*.

<sup>5</sup><http://download.geofabrik.de/europe.html>

Posteriormente as secções são intersectadas com os *buffers* exteriores e com a FGCI, e os polígonos resultantes da intersecção (Figura 4.5) são guardados para depois serem usados para extrair os dados das imagens satélite.

## 5.4 Extração de Dados

O primeiro passo da extração da informação das imagens satélite consiste no recorte das secções (*buffer* exterior e FGCI) para todas as bandas e índices de vegetação, em todas as datas. Este processo é feito para as imagens do Sentinel-1 e do Sentinel-2. De todas as fases da abordagem, incluindo as fases de treino e classificação, esta é a que tem os maiores tempos de processamento.

O recorte das imagens é feito recorrendo à ferramenta *rasterio*, usando a função *mask* que recebe como parâmetro os polígonos a serem usados para recorte. Apesar do recorte de uma imagem satélite usando um polígono ser relativamente rápido (média de 15 milissegundos), como existem milhares polígonos e várias imagens por data e o recorte foi feito sequencialmente polígono a polígono e imagem a imagem, o processo de recorte desta informação é lento.

Na Figura 5.10 pode ser visto um gráfico com os tempos de recorte (em segundos) para as imagens do Sentinel-1 e Sentinel-2 para os diferentes tipos de FGCI. O recorte das imagens do Sentinel-1 é consideravelmente mais rápido que do Sentinel-2 uma vez que para cada data apenas existem 2 imagens (VV e VH), já no caso do Sentinel-2 existem 13 imagens (bandas e índices de vegetação) por data.

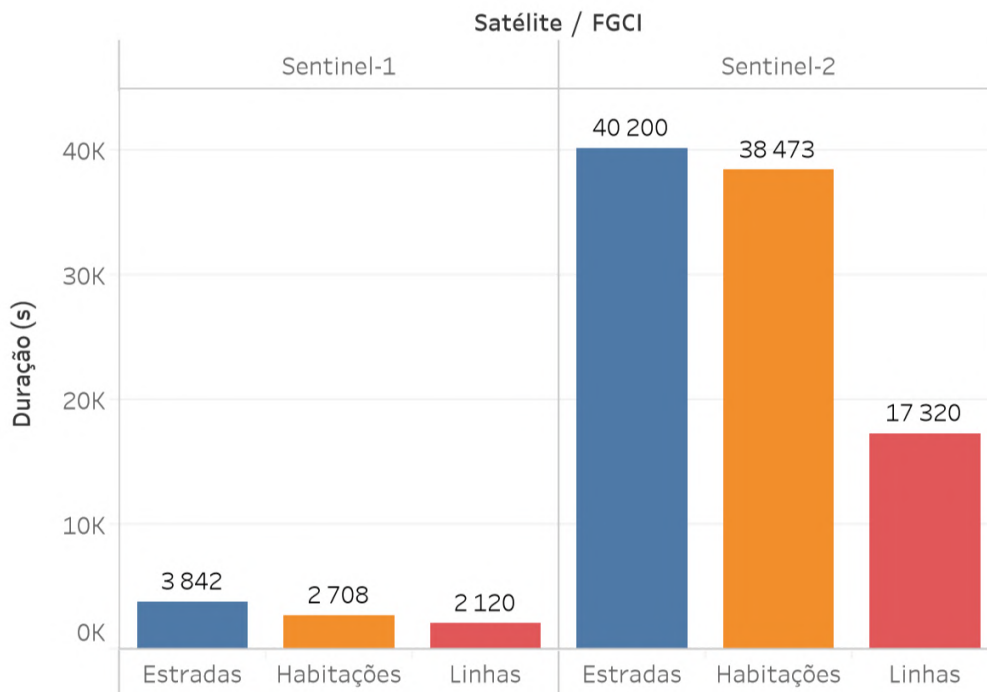


Figura 5.10: Tempos de recorte das imagens por FGCI e por satélite.

Cada imagem recortada é guardada num ficheiro com o formato `<secção>_<zona>_<banda ou índice>_<data>.tiff` (e.g. `secção1_buffer_NDVI_20180812.tiff`) resultando em duas imagens por secção (*buffer* e interior da FGCI), para cada data e para cada banda e índice de vegetação. Todas estas imagens são lidas posteriormente para determinar a média no interior da FGCI e no *buffer* para as bandas e índices, sendo o resultado agrupado por secção e guardado num ficheiro *csv*. Estes dados resultantes já podem ser analisados recorrendo a séries temporais ou algoritmos de aprendizagem automática.

### 5.5 Conclusões

A preparação dos dados para serem processados é um passo importante para poder ser realizada a análise das intervenções nas FGCI.

O pré-processamento da informação vetorial é composto por vários componentes importantes. A criação dos *buffers* exteriores e a divisão das faixas em secções são passos fundamentais para analisar as faixas usando a abordagem de comparação de vegetação entre o interior e exterior.

A automatização destes processos é essencial pois devido ao elevado número de polígonos que descrevem as faixas, não era concebível realizar estes processos de forma manual recorrendo a software de informação geográfica. No entanto, o custo temporal pode ser bastante elevado dependendo do número de imagens que está a ser usado.

Com este conjunto de métodos criados, toda a informação necessária para realizar a análise das FGCI já pode ser extraída e agrupada para depois alimentar os algoritmos de aprendizagem automática.



## TRABALHO EXPERIMENTAL

### 6.1 Introdução

Na fase experimental foi usada como informação base os dados extraídos das imagens dos satélites Sentinel-2 e Sentinel-1, através das secções criadas ao longo das faixas.

O trabalho experimental realizado foi dividido em duas partes: análise estática e análise temporal. Na análise estática recorreu-se a uma imagem de uma data após terem sido realizadas intervenções nas FGCI do concelho de Mação. Já na análise temporal foram usadas imagens de várias datas do ano 2018 e extraídas características dos padrões temporais destas imagens.

Testaram-se abordagens diferentes de classificação e comparados conjuntos de dados diferentes de forma a perceber qual a combinação de dados e algoritmos que melhor identifica as faixas que foram intervencionadas. No Anexo A encontram-se os resultados obtidos em todas as experiências da análise estática e temporal que recorreram a algoritmos de classificação.

### 6.2 Especificações de Hardware

Todas as experiências realizadas foram feitas usando um computador portátil com o sistema operativo Windows 10 e com as seguintes especificações de hardware:

- **CPU:** Intel Core i7-8750H Hexa Core, 2.20 GHz até 4.10 GHz
- **RAM:** 16GB DDR4 2666MHz
- **GPU:** NVIDIA GeForce GTX 1060 6GB GDDR5

### 6.3 Dados para Classificação

Para realizar a análise das faixas foram usadas 20 imagens do Sentinel-2 e 15 do Sentinel-1. Na Figura 6.1 encontra-se a distribuição destas imagens ao longo do ano 2018. Para o Sentinel-2 foram selecionadas imagens com uma percentagem de nuvens inferior a 10% e com o nível de produto L2A, que disponibiliza as imagens com correção atmosférica aplicada. Para o Sentinel-1 foram usadas imagens com o tipo de produto GRD, todas elas obtidas em órbita descendente e com os valores de intensidade convertidos para *Gamma0*.



Figura 6.1: Datas das imagens usadas de ambos os satélites.

Para cada uma das secções criadas ao longo das faixas extraíram-se os dados das imagens satélite recorrendo à informação vetorial das FGCI e dos *buffers* exteriores. Na tabela 6.1 está presente o número de secções por tipo de faixa, assim como a média da área de cada secção.

Tabela 6.1: Número de secções por tipo de FGCI e a média de área das secções.

FGCI	Nº de Secções Criadas	Nº de Secções Consideradas	Área Média das Secções
Estradas	5190	4205	4659 m <sup>2</sup> (0,47 ha)
Linhas Elétricas	2460	2293	4805 m <sup>2</sup> (0,48 ha)
Habitações	2994	2308	11276 m <sup>2</sup> (1,13 ha)

De notar que, uma vez que a abordagem adotada depende da informação dos *buffers* exteriores às faixas, e que estes apenas foram criados em zonas onde existe vegetação, as secções que não têm *buffer* exterior (e.g. secções de faixas ao longo de estradas que passam no meio de aglomerados habitacionais) não foram consideradas no momento da classificação.

Uma vez que são usadas imagens de várias datas e são extraídas métricas temporais destas imagens e tendo em conta que os conjuntos de dados usados podem ter dimensões diferentes, a dimensionalidade dos dados irá variar de acordo com a análise que está a ser feita (estática ou temporal) e com o conjunto de dados usado. Na Tabela 6.2 está representado o número total de características geradas a partir dos conjunto de dados apresentados na secção 4.4.



Tabela 6.2: Número de características de cada conjunto de dados para os tipos de análise realizados.

Conjunto de Dados	Análise Estática (Sentinel-2)	Análise Temporal (Sentinel-2)	Análise Temporal (Sentinel-1)	Análise Temporal (Sentinel-1 e 2)
CD_TUDO	52	260	80	340
CD_ÍNDICES	20	100	80	180
CD_BANDAS	32	160	80	240
CD_FUSÃO	32	160	80	240

As únicas características que foram variadas de acordo com o conjunto de dados foram as do Sentinel-2, para o Sentinel-1 foi sempre usado o mesmo conjunto de características (VV, VH, VV-VH e VV/VH), daí o número de características deste satélite ser igual para todos os conjuntos de dados.

## 6.4 Afinação de Parâmetros

Para construir modelos que permitem resolver um dado problema da melhor forma possível é necessário ajustar os parâmetros dos algoritmos de classificação.

A fim de encontrar os melhores valores dos parâmetros para os diferentes algoritmos de classificação usados, foi usado *cross-validation*. Porém, devido à quantidade de dados e aos diferentes conjuntos de dados usados, a afinação destes parâmetros para os diferentes algoritmos pode ser muito demorada, especialmente para os algoritmos RF e XGBoost que têm vários parâmetros de gamas diferentes.

Desta forma, ao invés de ter sido feita uma pesquisa exaustiva de valores para os diferentes parâmetros, foi usado o *RandomizedSearchCV* da biblioteca *scikit-learn*. Este método permite associar uma distribuição contínua aos parâmetros que aceitam valores contínuos e um conjunto de valores discretos para outros tipos de parâmetros, sendo que posteriormente são selecionados aleatoriamente um conjunto de configurações destes parâmetros. Assim, foi usada a distribuição uniforme para os parâmetros contínuos e um conjunto de 50 configurações de valores de parâmetros.

Foi realizada uma afinação de parâmetros para os diferentes conjuntos de dados, algoritmos e para os diferentes tipos de faixas. Foi usado o *RandomizedSearchCV* com 70% dos dados e com a opção de validação cruzada com 5 *folds* para fazer treino e validação, com os restantes 30% foi testado o melhor modelo.

Segue-se uma descrição dos parâmetros e os respetivos valores que foram usados para afinar cada um dos classificadores.

### 6.4.1 KNN

No algoritmo KNN foram usados 5 parâmetros: *weights* que permite definir a forma de atribuir o peso aos pontos vizinhos; *p* que faz variar a métrica de cálculo da distância

entre pontos (distância de *Minkowski*) ; *n\_neighbors* que define o número de vizinhos, de um ponto, a ter em conta; *algorithm* o algoritmo para computar os vizinhos mais próximos (e.g. *ball\_tree*, *kd\_tree*), *leaf\_size* usado para os algoritmos *ball\_tree* e *kd\_tree*.

Os conjuntos de valores usados para estes parâmetros estão presentes na Tabela 6.3.

Tabela 6.3: Valores usados para os parâmetros do KNN.

Parâmetro	Valores Testados
<i>weights</i>	{uniform, distance}
<i>p</i>	{1,2,3}
<i>n_neighbors</i>	{2,4,6,8,10,12,14}
<i>leaf_size</i>	{20,25,30,35,40,45}
<i>algorithm</i>	{auto, <i>ball_tree</i> , <i>kd_tree</i> }

#### 6.4.2 SVM

No algoritmo SVM foram usados 4 parâmetros: *kernel* função de *kernel* que pode ser radial (*rbf*), linear, polinomial (*poly*) ou sigmoide (*sigmoid*); *C* que penaliza o erro (i.e. valores baixos maior margem entre o hiperplano e os vetores de suporte, valores altos menor margem); *degree* define o grau da função de *kernel* polinomial e *gamma* coeficiente das funções *rbf*, *poly* e *sigmoid*.

Na Tabela 6.4 estão presentes os valores que foram utilizados para afinar este classificador.

Tabela 6.4: Valores usados para os parâmetros do SVM.

Parâmetro	Valores Testados
<i>C</i>	uniforme(0.001, 600)
<i>degree</i>	{3,4,5,6,7}
<i>gamma</i>	uniforme(0.001,600)
<i>kernel</i>	{ <i>rbf</i> ,linear, <i>poly</i> , <i>sigmoid</i> }

#### 6.4.3 RF

No algoritmo RF foram usados 6 parâmetros: *criterion* função que mede a qualidade da divisão dos nós das árvores de decisão; *max\_depth* que define profundidade máxima de uma árvore; *max\_features*, estabelece o número de características a considerar quando for realizado uma divisão de nós; *min\_samples\_leaf* número mínimo de amostras necessárias para um nó ser considerado uma folha; *min\_samples\_split*, número mínimo de amostras necessárias para realizar uma divisão do nó e *n\_estimators* número de árvores a serem criadas.

Na Tabela 6.5 podem ser consultados os valores que foram usados para afinar este classificador.

Tabela 6.5: Valores usados para os parâmetros do RF.

Parâmetro	Valores Testados
<i>criterion</i>	{ <i>gini</i> , <i>entropy</i> }
<i>max_depth</i>	{8,10,12,14,16,18,20,22,24}
<i>max_features</i>	uniforme(0.05, 0.94)
<i>min_samples_leaf</i>	{1,2,3,4,5}
<i>min_samples_split</i>	{2,3,4,5,6}
<i>n_estimators</i>	{100,200,300,400,500,600}

#### 6.4.4 XGBoost

No algoritmo XGBoost foram usados 8 parâmetros: *colsample\_bytree* indica a percentagem de colunas a serem usadas para construir cada árvore; *gamma* o valor mínimo para um nó ser dividido; *learning\_rate* que é usado para reduzir os pesos das características entre os passos de *boosting*, tornando o processo mais conservador; *max\_depth* estabelece a profundidade máxima de uma árvore; *n\_estimators* o número de árvores a serem criadas; *reg\_alpha*, termo de regularização dos pesos das amostras; *scale\_pos\_weight* que controla o equilíbrio entre pesos positivos e negativos e *subsample* a percentagem de instâncias de treino a serem usadas.

Os valores usados para cada um destes parâmetros podem ser consultados na Tabela 6.6.

Tabela 6.6: Valores usados para os parâmetros do XGBoost.

Parâmetro	Valores Testados
<i>colsample_bytree</i>	uniforme(0.05,0.94)
<i>gamma</i>	uniforme(0.1, 10)
<i>learning_rate</i>	uniforme(0.001, 0.65)
<i>max_depth</i>	{2,4,6,8,10,12,14,16}
<i>n_estimators</i>	{200,300,400,500,600}
<i>reg_alpha</i>	uniforme(0.01, 1.0)
<i>scale_pos_weight</i>	uniforme(0,2)
<i>subsample</i>	uniforme(0.01, 1.0)

## 6.5 Análise Estática

No ano de 2018 foram intervencionadas várias faixas ao longo de estradas e linhas elétricas no concelho analisado, Mação. Estas faixas foram intervencionadas em alturas diferentes sendo que a primeira intervenção que pôde ser detetada através da visualização das imagens satélite ocorreu entre 25/04/2018 e 15/05/2018. Nas Figuras 6.2a e 6.2b está presente o NDVI calculado nestas duas datas, como podemos observar existe uma mudança significativa na vegetação ao longo de uma estrada que faz parte da rede de faixas.

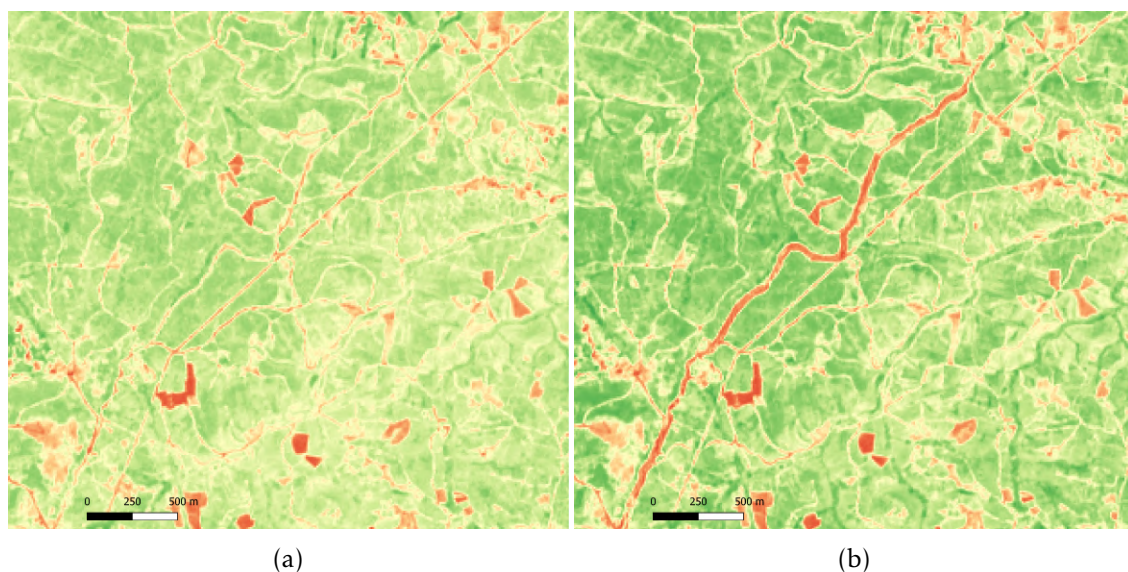


Figura 6.2: (a) NDVI calculado usando imagens do Sentinel-2 de 25-04-2018. (b) NDVI calculado usando imagens do Sentinel-2 de 15-05-2018.

Entre os meses de Abril e Agosto ocorreram as operações de limpeza em grande parte das faixas que foram intervencionadas no ano 2018. Como tal, foi usada uma imagem do início do mês de Setembro, após as intervenções terem sido realizadas, para efetuar uma primeira análise recorrendo a algoritmos de aprendizagem supervisionada e não supervisionada.

### 6.5.1 Agrupamento

As primeiras experiências realizadas usando os dados das imagens satélite foi recorrendo ao algoritmo de aprendizagem não supervisionada *K-Means*.

O *K-Means* é um algoritmo de agrupamento que particiona um conjunto de pontos num número predefinido de partições, que é dado pelo parâmetro  $K$ . Foi usado este algoritmo de aprendizagem não supervisionada, uma vez que quando foram realizadas estas experiências os dados oficiais das faixas intervencionadas neste concelho ainda não estavam disponíveis, não sendo possível recorrer a algoritmos de classificação.

Foram realizadas duas experiências com este algoritmo que são apresentadas nesta secção: a primeira em faixas ao longo de estradas e a segunda em faixas ao redor de habitações.

#### 6.5.1.1 Faixa ao Longo de Estradas

A primeira análise realizada foi feita através do algoritmo de agrupamento *K-Means* e em faixas ao longo de estradas, numa zona no norte do concelho de Mação.

A análise consistiu em usar o índice de vegetação NDVI e o valor deste dentro da FGCI, no *buffer* exterior e a diferença entre ambos ( $diff_{NDVI}$ ), de modo a tentar perceber se existe

uma relação entre os *clusters* gerados pelo *K-Means* e a presença de vegetação nas FGCI ao longo das estradas. Foi usado o parâmetro  $K = 3$ , de forma a tentar estabelecer uma relação entre os dados do NDVI e secções que aparentavam ter sido intervencionadas, não intervencionadas ou zonas que podiam ou não ter sido intervencionadas.

Na Figura 6.3a pode ser vista a área que foi analisada e as secções ao longo das FGCI das estradas que foram analisadas. Das 4205 secções das faixas ao longo das estradas foram analisadas 275, a branco estão as secções que correspondem às que foram intervencionadas e a vermelho às que não foram.

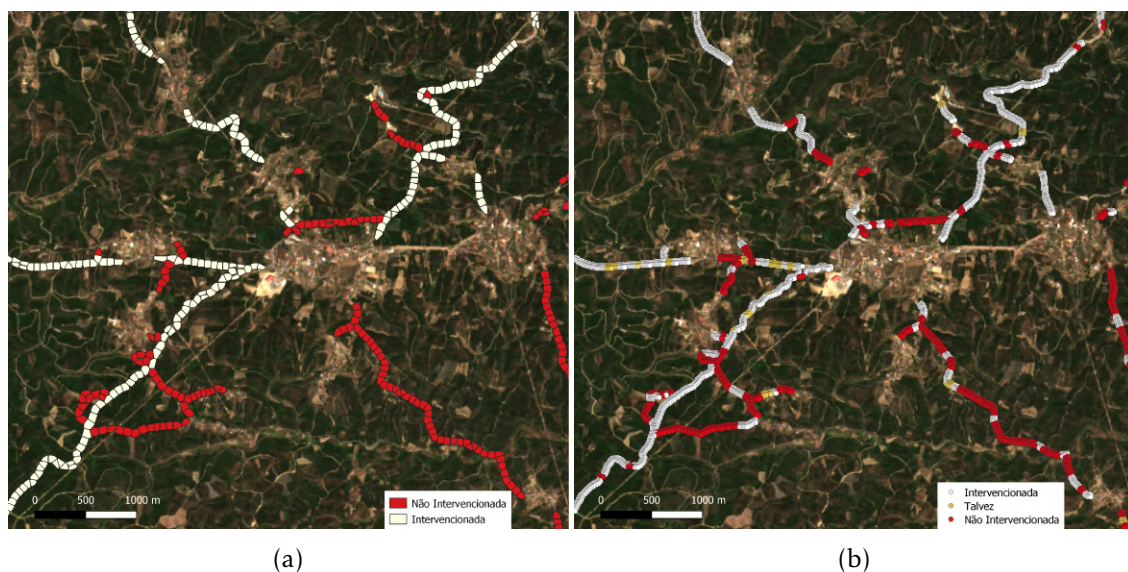


Figura 6.3: (a) Secções consideradas para realizar a classificação. (b) Secções que foram intervencionadas (branco) e não intervencionadas (vermelho).

O resultado de executar o algoritmo *K-Means* com os dados referidos a cima pode ser visto na Figura 6.3b. Cada ponto representado na figura representa um píxel, sendo que estes estão agrupados de acordo com as secções da Figura 6.3a. Podemos observar que a maioria das secções intervencionadas foram corretamente classificadas, já as não intervencionadas existe algumas que foram incorretamente classificadas como intervencionadas.

Uma vez que posteriormente foi obtida informação precisa sobre as faixas intervencionadas, foi possível calcular algumas métricas sobre estes resultados e construir uma matriz de confusão (Tabela 6.7). Como a classe de referência é binária, não foi possível calcular as métricas da classe *Talvez*.

Apenas 17 das 275 secções foram atribuídas à classe *Talvez*, sendo que a maioria destas encontra-se em faixas que foram intervencionadas. Das 156 secções classificadas como intervencionadas apenas 31 foram incorretamente classificadas, resultando num F1-score de 0,8 enquanto que as secções não intervencionadas obtiveram 0,75.

Os resultados deste agrupamento foram bastante razoáveis e mostraram que no caso de falta de dados para realizar o treino de classificadores, o uso de algoritmos de aprendizagem não supervisionada pode gerar resultados satisfatórios.



Tabela 6.7: Matriz de confusão dos resultados do agrupamento realizado.

		Referência		Total	Precisão	Recall	F1-score
		Interven.	Não Interven.				
Previsto	Interven.	125	31	156	0,80	0,79	0,80
	Não Interven.	19	83	102	0,81	0,70	0,75
	Talvez	13	4	17	-	-	-
	Total	157	118				

### 6.5.1.2 Faixas ao Redor de Habitações

Tendo por base a experiência anterior e os seus resultados, foi realizada a mesma experiência com o mesmo conjunto de dados, mas aplicada a faixas ao redor de habitações.

Como é usado o NDVI dentro da faixa, no *buffer* exterior e a  $dif_{NDVI}$  só foram tidas em conta as secções que têm *buffer* exterior, logo só as secções que se encontram na margem exterior da faixa é que foram consideradas.

Uma vez que não faz parte dos dados disponibilizados pela câmara de Mação faixas ao redor de habitações que tenham sido intervencionadas em 2018, não foi possível calcular métricas que permitissem estimar a qualidade do agrupamento resultante. Tal como no agrupamento realizado nas faixas ao longo de estradas, foi usado o *K-Means* com  $K = 3$ , para tentar estabelecer uma relação com secções possivelmente intervencionadas, não intervencionadas ou que possam ou não ter sido intervencionadas.

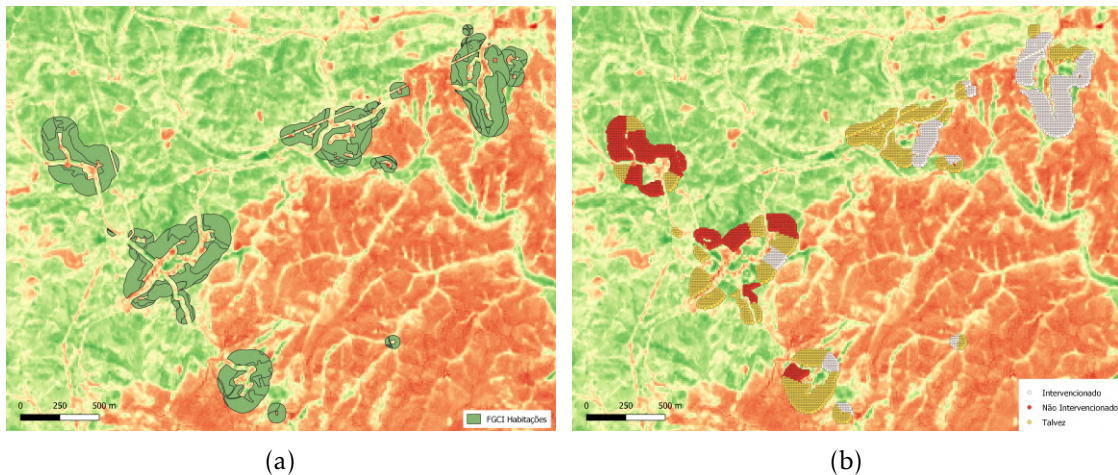


Figura 6.4: (a) Faixas faixas ao redor de habitações na zona analisada. (b) Resultado do agrupamento das secções de faixas ao redor de habitações.

Foi escolhida uma zona que foi parcialmente afetada por incêndios em 2017 e usada uma imagem do Sentinel-2 de 29 de Julho de 2018 para a qual foi calculado o NDVI (Figura 6.4a). Podemos ver em tons de vermelho a zona que foi afetada pelos incêndios e que tem o NDVI baixo, enquanto que em tons de verde zonas de vegetação que não foi afetada.

Na Figura 6.4b podemos ver o resultado do agrupamento. Pode ser observado que nas faixas que se encontram dentro da zona onde os valores do NDVI são bastante baixos, as secções foram todas agrupadas como sendo intervencionadas ou como talvez intervencionadas. Já nas zonas onde o NDVI é mais elevado e que não foram afetadas pelos incêndios temos secções agrupadas como não intervencionadas e em alguns casos como talvez intervencionadas.

Esta análise é uma análise muito superficial pois não existem dados que corroborem o que está a ser observado, no entanto tendo por base os resultados do agrupamento nas faixas ao longo de estradas, esta pode ser uma abordagem a ter em conta quando dados oficiais do terreno não se encontram disponíveis.

### 6.5.2 Classificação Supervisionada

Através da informação disponibilizada por Mação foi possível recorrer a algoritmos de aprendizagem supervisionada para realizar uma classificação das faixas que foram intervencionadas.

Apenas foram analisadas faixas ao longo das estradas e linhas elétricas, uma vez que estas foram as faixas para as quais foram obtidos dados oficiais sobre as intervenções.

Foram usados todos os conjuntos de dados presentes na Tabela 4.2 com informação do satélite Sentinel-2 e analisada a performance dos algoritmos KNN, SVM, RF e XGBoost em cada um deles.

#### 6.5.2.1 Faixas ao Longo de Estradas

Nas faixas ao longo de estradas foram testados os diferentes conjuntos de dados usando os dados do Sentinel-2.

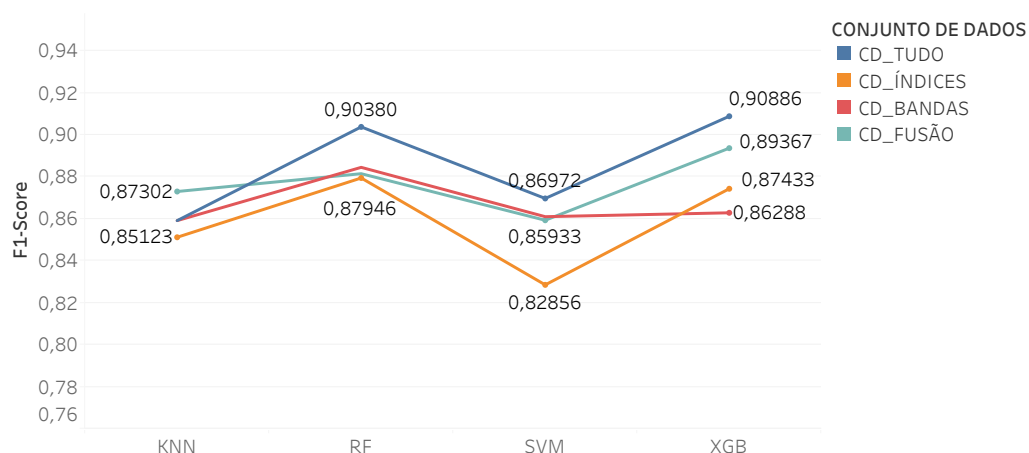


Figura 6.5: F1-score para os diferentes algoritmos e conjuntos de dados nas faixas ao longo das estradas.

Na Figura 6.5 estão presentes os valores da métrica F1-score e na Figura 6.6 da métrica Kappa, para os algoritmos usados e para os diferentes conjuntos de dados. Dos algoritmos

escolhidos o RF e o XGBoost foram aqueles que se destacaram, com o XGBoost a obter resultados ligeiramente melhores.

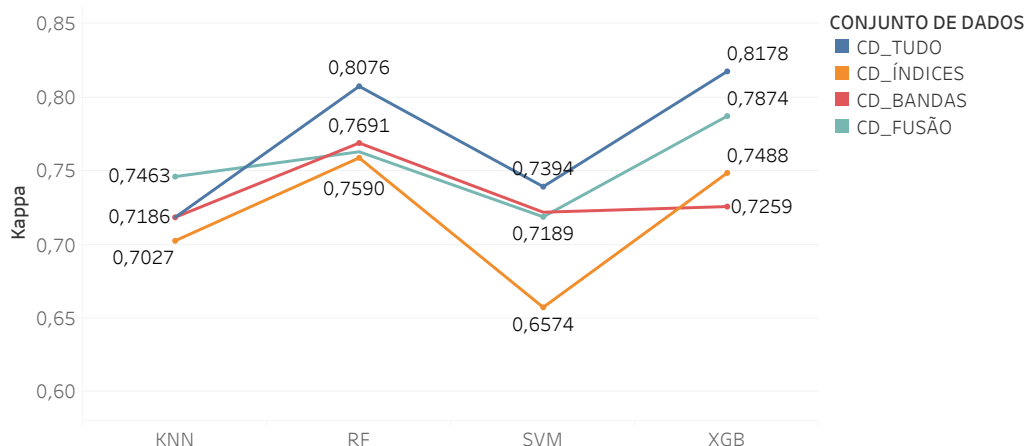


Figura 6.6: Kappa para os diferentes algoritmos e conjuntos de dados nas faixas ao longo das estradas.

Em relação aos conjuntos de dados podemos observar que em todos os algoritmos, com exceção do KNN, o conjunto de dados com o qual foram obtidos os melhores valores do F1-score e do Kappa foi o *CD\_TUDO*. A escolha do conjunto de dados mostrou ser determinante na qualidade dos resultados, uma vez que em alguns casos fez uma diferença superior a 4% no F1-score.

O pior resultado foi obtido usando o conjunto de dados *CD\_ÍNDICES* e o algoritmo SVM, com um F1-score de 0,83 e com o Kappa 16% mais baixo com 0,66.

Ao analisar mais detalhadamente, através de uma matriz de confusão (Tabela 6.8), a combinação que obteve melhores resultados (XGBoost usando o conjunto de dados *CD\_TUDO*) podemos observar que a classe das faixas não intervencionadas foi a que obteve melhores resultados, com 0,99 de F1-score e precisão e com *recall* de 0,98 e 0,99, respetivamente. As secções intervencionadas tiveram valores mais baixos com um F1-score de 0,83.

Tabela 6.8: Matriz de confusão para o algoritmo XGBoost usando o conjunto de dados *CD\_TUDO*.

		Referência		Total	Precisão	Recall	F1-score
		Interven.	Não Interven.				
Previsto	Interven.	89	12	101	0,88	0,79	0,83
	Não Interven.	24	1263	1287	0,98	0,99	0,99
	Total	113	1275				
	F1-Score Global	0,91					
Kappa Global		0,82					



### 6.5.2.2 Faixas ao Longo de Linhas Elétricas

As faixas ao longo das linhas elétricas foram analisadas de forma idêntica às faixas ao longo das estradas. Novamente os algoritmos com melhores resultados foram o RF e o XGBoost (Figura 6.7 e 6.8), com o RF obter resultados ligeiramente superiores.

Para estas faixas o algoritmo que obteve piores resultados foi o SVM, tal como aconteceu na experiência realizada nas FGCI das estradas.

A diferença de performance entre os diferentes conjuntos dados foi mais notada no algoritmo RF em que o conjunto de dados *CD\_ÍNDICES* obteve um F1-score de 0,80, aproximadamente 4% inferior ao *CD\_TUDO* que foi o melhor. O KNN foi o que mostrou variar menos com os conjuntos de dados, de tal maneira que a diferença entre o melhor e o pior foi inferior a 1%.

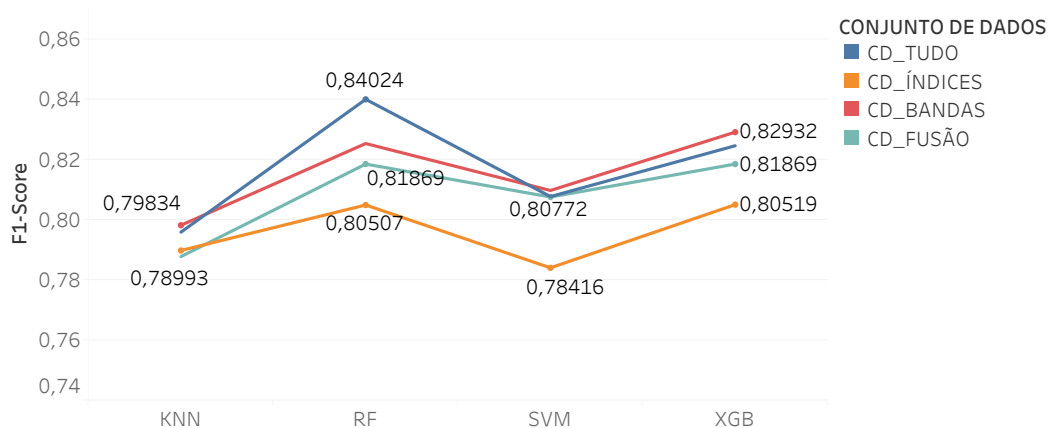


Figura 6.7: F1-score para os diferentes algoritmos e conjuntos de dados, nas faixas ao longo das linhas elétricas.

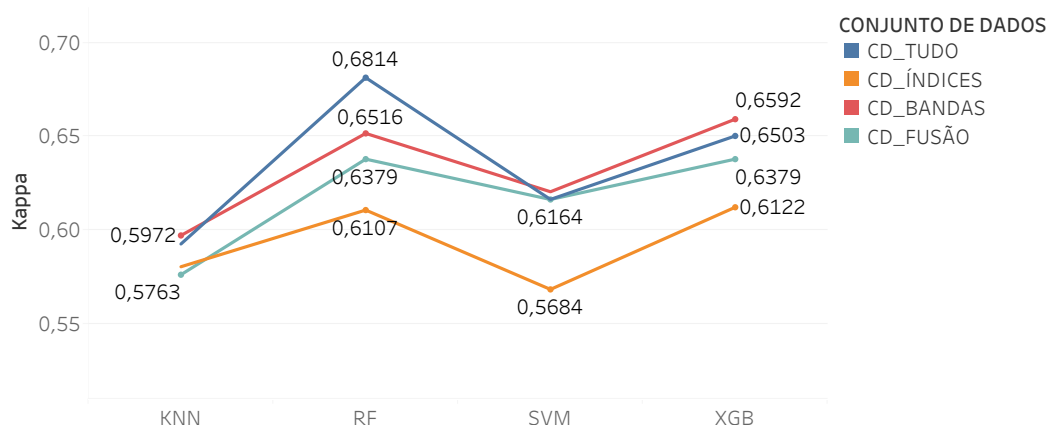


Figura 6.8: Kappa para os diferentes algoritmos e conjuntos de dados, nas faixas ao longo das linhas elétricas.

De notar também que os conjuntos de dados *CD\_TUDO* e *CD\_BANDAS* foram os conjuntos que obtiveram os melhores resultados para todos os algoritmos, em alguns

casos obtendo valores muito próximos um do outro.

Na matriz de confusão presente na tabela 6.9 podemos ver os resultados discriminados por classe, para o algoritmo e conjunto de dados que obteve os melhores resultados: o RF usando o conjunto de dados *CD\_TUDO*.

Tabela 6.9: Matriz de confusão para o algoritmo RF usando o conjunto de dados *CD\_TUDO*.

		Referência					
		Interven.	Não Interven.	Total	Precisão	Recall	F1-score
Previsto	Interven.	125	25	150	0,83	0,68	0,75
	Não Interven.	58	549	607	0,90	0,96	0,93
	Total	183	574				
	F1-Score Global Kappa Global	0,84 0,68					

Entre as duas classes as métricas mais elevadas foram da classe das secções não intervencionadas, com um F1-score de 0,93. A classe das secções intervencionadas teve resultados mais baixos, com F1-score de 0,75. Globalmente o F1-score foi de 0,84 e o kappa consideravelmente inferior, com 0,68 mostrando que em parte dos resultados a concordância entre previsão e o dados de referência pode ter sido por acaso.

No geral, as classificações das intervenções nas faixas ao longo de linhas elétricas foram inferiores às das estradas, com o F1-score global a ser 7% inferior.

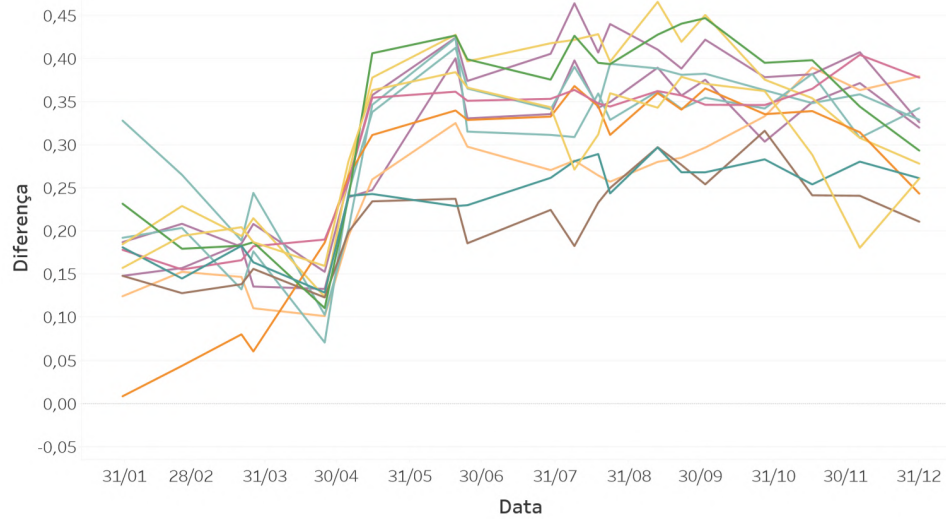
## 6.6 Análise Temporal

A maioria das experiências realizadas recorreram a séries temporais de índices e das bandas dos satélites.

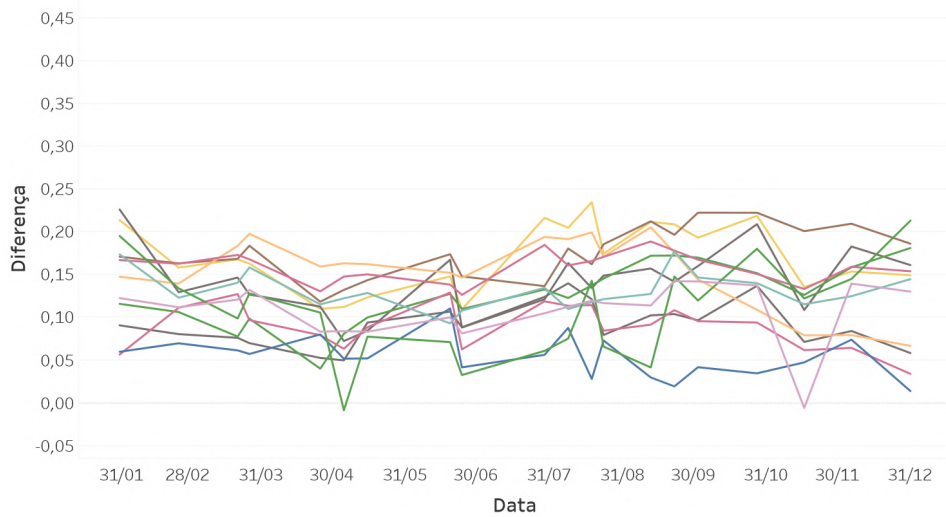
O sinal temporal das múltiplas características pode ser indicador do comportamento da vegetação numa determinada zona ao longo do tempo. Ao usar indicadores como a diferença ou o rácio de vegetação entre os *buffers* exteriores e o interior das FGCI, mudanças drásticas da vegetação no exterior ou no interior podem ser facilmente notadas no sinal destes indicadores.

Como exemplo deste comportamento, pode ser observado na Figura 6.9a o comportamento da diferença do NDVI ( $dif_{NDVI}$ ) entre o *buffer* exterior e o interior de algumas secções de uma faixa ao longo de uma estrada que foi intervencionadas no ano 2018. Já na Figura 6.9b o comportamento do  $dif_{NDVI}$  mas em secções não intervencionadas.

Como podemos observar, no momento onde ocorre a intervenção da faixa, a  $dif_{NDVI}$  das secções apresentadas sobe para valores que em alguns casos aproximam-se dos 0,45. Enquanto que nas secções das faixas que não foram sujeitas a nenhum tipo de intervenção os valores da  $dif_{NDVI}$  mantém-se aproximadamente constantes ao longo do ano.



(a)



(b)

Figura 6.9: Comportamento da  $dif_{NDVI}$  ao longo de 2018 em secções que foram intervencionadas entre 25-04-2018 e 15-05-2018 (a) e em secções que não foram intervencionadas (b).

Este tipo de sinal pode ser usado para classificar as secções, e uma forma simples de o fazer é estabelecendo um limite, e se o sinal for superior a esse limite pode ser indicador de que houve uma intervenção. Outra abordagem é extrair métricas como a média, máximo, mínimo, etc, destes padrões temporais e inserir esses valores em algoritmos de aprendizagem automática.

Nesta secção são analisadas estas abordagens usando dados do Sentinel-2, Sentinel-1 e a junção de ambos.

### 6.6.1 Classificação Através de Imposição de Limite

A primeira abordagem usada para determinar se uma determinada secção foi ou não intervencionada foi recorrendo às séries temporais de  $dif_x$  dos índices de vegetação NDVI, NDWI, SAVI, IRECI e fazer variar um limite ( $l$ ). Quando  $dif_x \geq l$  a secção é classificada como intervencionada e caso contrário não intervencionada.

Apenas foram considerados valores de  $l$  superiores a zero uma vez que quando  $dif_x > 0$  indica que a vegetação no *buffer* exterior à faixa é mais vigorosa (ou que tem um maior conteúdo de água, no caso do NDWI) em relação ao interior. Como tal, foram usados valores de  $l$  que variam entre 0,04 e 0,65 em intervalos de 0,02.

Os resultados foram comparados com os dados disponibilizados pela Câmara Municipal de Mação e os índices que obtiveram os melhores resultados na classificação de secções intervencionadas foram o NDVI e o SAVI, com um *F1-score* de 0,69 e 0,51 respetivamente. Para o NDVI o limite que obteve o melhores resultados foi de 0,28 e para o SAVI foi de 0,14.

O comportamento da precisão, *recall* e *F1-score* para o índice NDVI pode ser observado na Figura 6.10. Nas secções intervencionadas (Figura 6.10a) podemos observar uma subida da precisão até valores de  $l$  entre os 0,4 e 0,5, sendo que entre 0,5 e 0,6 todas as métricas atingem o valor 0, indicando que não existe nenhuma secção intervencionada em que o  $dif_{NDVI}$  atinja 0,6. O *F1-score* máximo é obtido para  $l = 0,28$ , próximo do cruzamento da curva do *recall* com a da precisão.

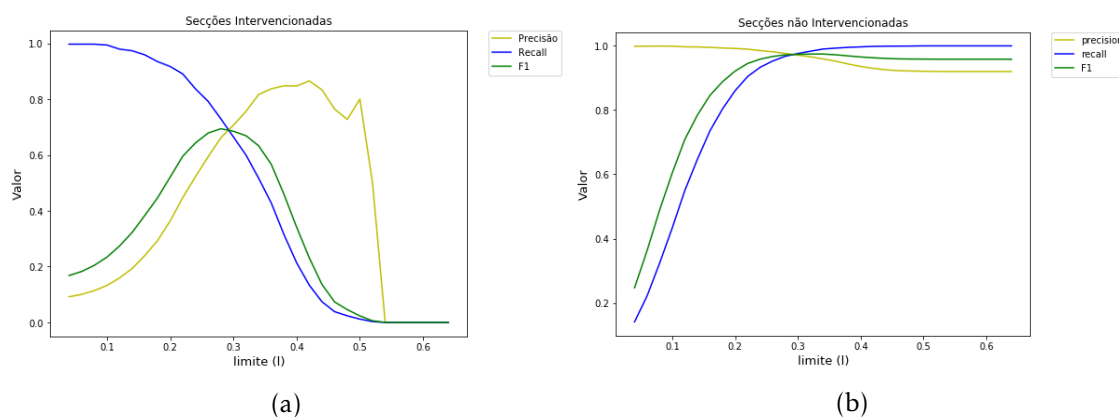


Figura 6.10: Valor das métricas *F1-score*, *recall* e precisão para os diferentes valores do limite, na classificação de secções intervencionadas (a) e não intervencionadas (b) usando o índice NDVI.

Na classificação de secções não intervencionadas (Figura 6.10b) foram observados melhores resultados. À medida que  $l$  aumenta a precisão também aumenta e o *recall* mantém-se constante até aproximadamente o valor do limite onde foram obtidos os melhores resultados para as secções intervencionadas ( $l = 0,28$ ), sendo que a partir daí o *recall* começa a descer. Para estas secções o *F1-score* máximo obtido foi de 0,97 usando

$l = 0, 32$ .

Embora esta seja uma abordagem simples mostrou bons resultados na classificação das secções não intervencionadas, já no caso das secções intervencionadas os resultados não foram tão bons, mas ainda assim foram razoáveis e mostram que este método pode ser aplicado para detetar mudanças na vegetação através da comparação de vegetação entre duas zonas (neste caso *buffer* e interior da FGCI).

## 6.6.2 Classificação Supervisionada

Com os algoritmos de classificação já mencionados, foram usados os padrões temporais das várias características e conjuntos de dados para tentar determinar se uma dada secção das faixas foi intervencionada.

Como mencionado na secção 4.6 foram extraídas características destes padrões temporais, e os seus valores foram inseridos nos algoritmos de aprendizagem automática.

Tal como na análise estática, foram comparadas as performances dos diferentes algoritmos e conjuntos de dados, no entanto também foram realizados testes com os padrões temporais do Sentinel-1 e com os padrões temporais do Sentinel-1 e Sentinel-2 juntos.

### 6.6.2.1 Faixas ao Longo de Estradas

As faixas ao longo das estradas foram analisadas usando informação do satélite Sentinel-2, com os diferentes conjuntos de dados criados, com informação do Sentinel-1 e usando a junção dos dados de ambos os satélites.

#### Sentinel-2

Todos os conjuntos de dados, com exceção do *CD\_ÍNDICES*, obtiveram resultados semelhantes. Comparativamente aos restantes conjuntos de dados o *CD\_ÍNDICES*, para os algoritmos KNN e SVM obteve resultados mais baixos, que chegaram a ter uma diferença de 23% em relação ao melhor conjunto de dados no caso do KNN.

Ao contrário do que aconteceu na abordagem estática, os resultados dos diferentes algoritmos foram bastante próximos (excluindo o conjunto de dados o *CD\_ÍNDICES*). Neste caso o melhor algoritmo foi o SVM obtendo um F1-score de 0,94, contudo o XGBoost e o RF tiveram resultados praticamente iguais.

Na Tabela 6.10 está presente a matriz de confusão do algoritmo SVM usando o conjunto de dados *CD\_TUDO*. A classe das secções não intervencionadas obteve métricas bastante elevadas próximas de 1, já no caso da classe das secções intervencionadas as métricas foram mais baixas, ainda assim boas, obtendo um F1-score de 0,89, com apenas 12 secções classificadas erradamente como intervencionadas.

O recurso a séries temporais revelou gerar bons resultados na classificação das secções intervencionadas e não intervencionadas pertencentes a faixas ao longo de estradas, mostrando uma melhoria significativa em relação à análise estática, com o F1-score da classe das secções intervencionadas a subir de 0,83 para 0,89.

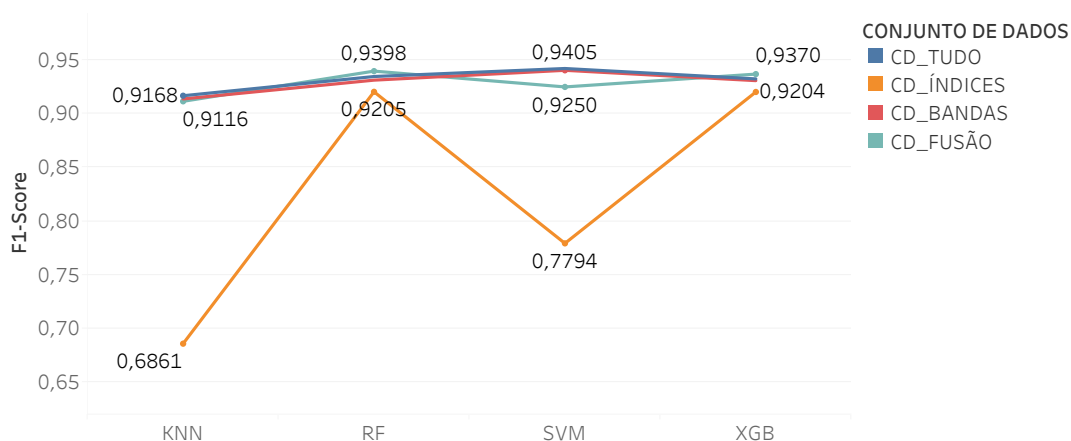


Figura 6.11: Valor da métrica F1-score para os diferentes algoritmos e conjuntos de dados com dados do Sentinel-2.

Tabela 6.10: Matriz de confusão para o algoritmo SVM usando o conjunto de dados *CD\_TUDO*.

		Referência		Total	Precisão	Recall	F1-score
		Interven.	Não Interven.				
Previsto	Interven.	101	12	113	0,89	0,89	0,89
	Não Interven.	12	1263	1275	0,99	0,99	0,99
	Total	113	1275				
	F1-Score Global Kappa Global	0,94 0,88					



Figura 6.12: Exemplos de falsos dos positivos registados com o SVM e conjunto de dados *CD\_TUDO*.

Nos exemplos da Figura 6.12 podemos observar duas secções que não foram interven-  
cionadas e que geraram falsos positivos. É possível observar que estas encontram-se em  
zonas de limite entre faixas não intervencionadas e faixas intervencionadas (Figura 6.12a)

ou no limite entre o exterior e o interior de aglomerados habitacionais (Figura 6.12b), fazendo com que parte da secção intersecte uma zona que foi intervencionada ou que não tem vegetação, resultando na classificação desta secção como intervencionada.

### Sentinel-1

Para este satélite foram usadas as métricas mencionadas na secção 4.6, extraídas dos padrões temporais das polarizações VV e VH, e a diferença e o rácio entre elas. O desempenho dos vários algoritmos usando estes dados pode ser observado na Figura 6.13.

As medidas de precisão para os dados deste satélite foram consideravelmente mais baixas que para o Sentinel-2, especialmente na métrica Kappa, ainda assim o RF obteve um F1-score de 0.67. Por outro lado, o SVM e KNN obtiveram resultados mais baixos chegando a métrica kappa a ser 0 no caso do SVM, uma vez que não conseguiram classificar corretamente secções intervencionadas, apenas as não intervencionadas.

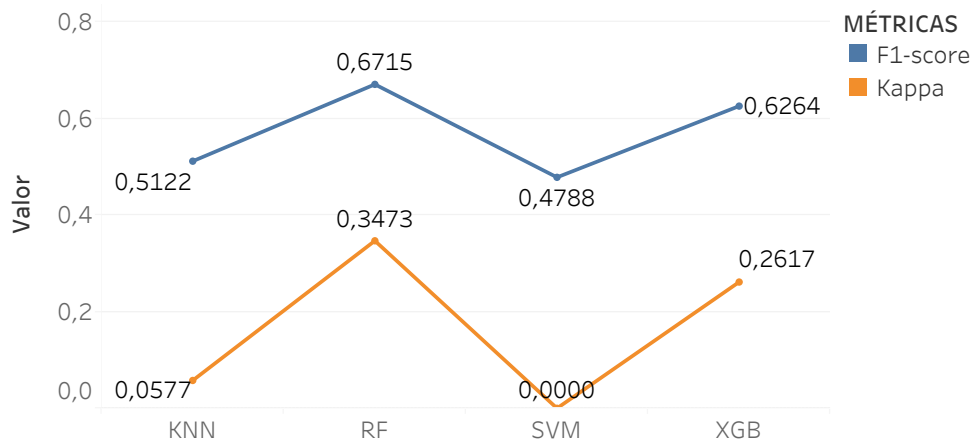


Figura 6.13: Valor das métricas F1-score e Kappa para os diferentes algoritmos para os dados do Sentinel-1.

Tabela 6.11: Matriz de confusão para o algoritmo RF para os dados do Sentinel-1.

		Referência		Total	Precisão	Recall	F1-score
		Interven.	Não Interven.				
Previsto	Interven.	33	26	59	0,56	0,29	0,38
	Não Interven.	80	1249	1329	0,94	0,98	0,96
	Total	113	1275				
	F1-Score Global	0,67					
	Kappa Global	0,34					

Ao analisar com mais detalhe a matriz de confusão (Tabela 6.11) dos resultados do RF podemos ver que as métricas das secções não intervencionadas tiveram uma grande influência no valor do F1-score global, uma vez que as secções intervencionadas apenas

tiveram um F1-score de 0,38 e com um *recall* baixo de 0.29.

Por outro lado, a classificação de secções não intervencionadas obteve muito bons resultados com um F1-score de 0.96, o que mostra que este tipo de dados pode ajudar a melhorar a classificação destas secções.

### Ambos os Satélites

A informação de ambos os satélites foi unida com a finalidade de perceber se a combinação destes dois satélites, que fornecem diferentes de características, pode ter um impacto positivo nos resultados.

Na Figura 6.14 encontram-se os valores do F1-score para os diferentes algoritmos e conjuntos de dados. No geral os valores das métricas mantiveram-se semelhantes aos valores obtidos usando apenas os dados do Sentinel-2. No caso do algoritmo SVM não houve alteração nos valores do F1-score obtidos, com exceção para o conjunto de *CD\_ÍNDICES* que piorou o seu F1-score em 1%.

No RF houve um decréscimo das métricas em todos conjuntos de dados e no XGBoost houve um aumento bastante reduzido nos conjuntos de dados *CD\_TUDO* e uma redução nas métricas do conjunto *CD\_BANDAS*.

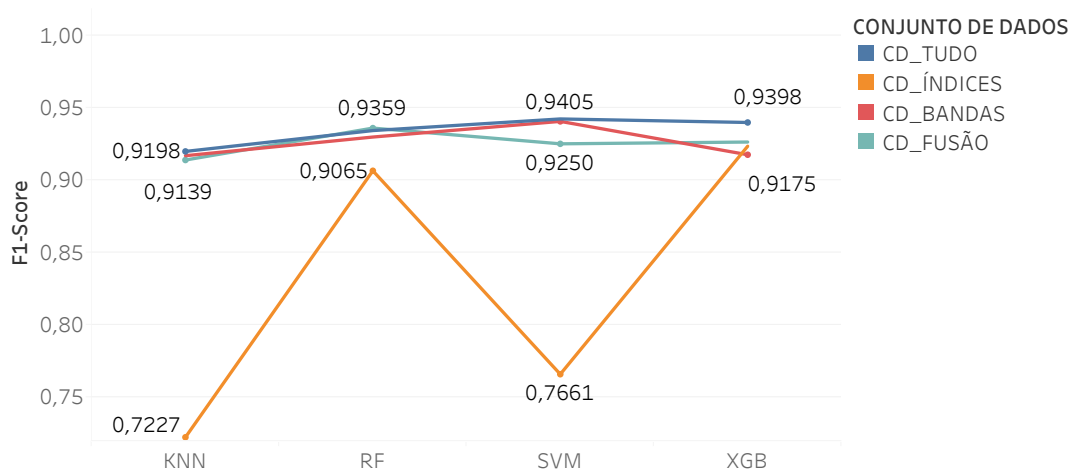


Figura 6.14: Valor da métrica F1-score para os diferentes algoritmos e conjuntos de dados, com dados do Sentinel-1 e Sentinel-2.

O maior impacto que esta união de dados teve foi no algoritmo KNN para o conjunto de dados *CD\_ÍNDICES* que melhorou o valor do F1-score em mais de 3%, passando de 0.69 para 0.72.

O SVM continuou a ser o melhor algoritmo com o mesmo F1-score que o observado na experiência com os dados do Sentinel-2, sendo que a matriz de confusão foi exatamente igual à matriz 6.10.



### 6.6.2.2 Faixas ao Longo de Linhas Elétricas

Nas faixas ao longo de linhas elétricas foi aplicada a mesma análise realizada às faixas ao longo de estradas, comparando os dados do Sentinel-2 e Sentinel-1 para os diferentes conjuntos de dados e algoritmos.

#### Sentinel-2

Os resultados para este tipo de faixa, usando os dados do Sentinel-2, foram inferiores aos das faixas ao longo das estradas. Tal como em experiências anteriores os algoritmos que se destacaram foram o XGBoost e o RF, com o XGBoost a ser melhor por uma margem tangencial.

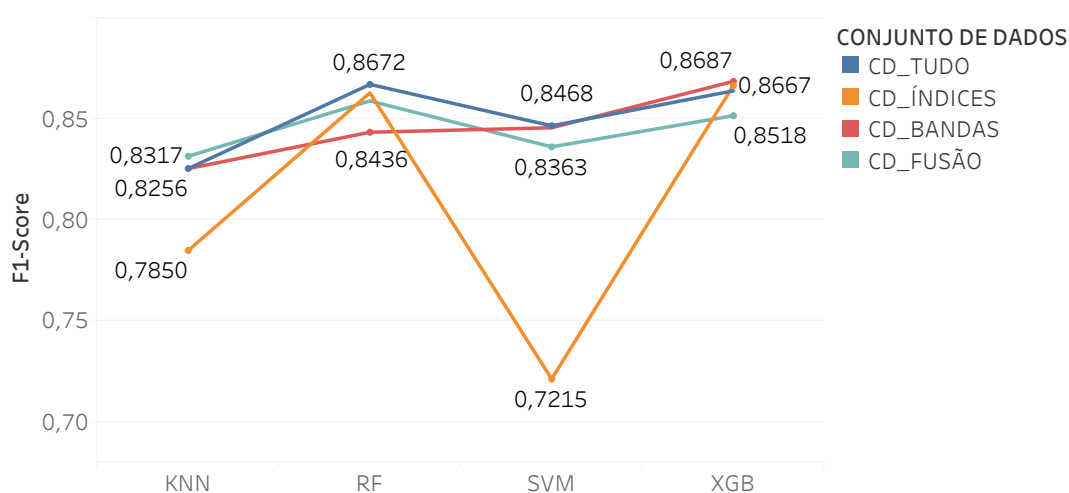


Figura 6.15: Valor da métrica F1-score para os diferentes algoritmos e conjuntos de dados.

Nesta experiência não houve nenhum conjunto de dados a se destacar pela positiva, uma vez que em todos os algoritmos foi um conjunto de dados diferente a obter os melhores resultados. Os piores resultados foram obtidos pelo SVM e KNN para o conjunto de dados *CD\_ÍNDICES*, no entanto este conjunto obteve resultados próximos do melhor F1-score para o algoritmo XGBoost e para o RF, mostrando que a performance deste conjunto de dados varia dependendo do algoritmo usado.

Tabela 6.12: Matriz de confusão para o algoritmo XGBoost usando o conjunto de dados *CD\_BANDAS*.

		Referência		Total	Precisão	Recall	F1-score
		Interven.	Não Interven.				
Previsto	Interven.	129	13	142	0,91	0,71	0,79
	Não Interven.	54	561	574	0,91	0,98	0,94
	Total	183	574				
	F1-Score Global	0,87					
		Kappa Global	0,74				

Na Tabela 6.12 podemos ver a matriz de confusão resultante do algoritmo XGBoost para o conjunto de dados que obteve os melhores resultados, o *CD\_BANDAS*. A classe das secções não intervencionadas foi a que obteve melhores resultados com um F1-score de 0,94, sendo que a classe das secções intervencionadas obteve um F1-score de 0,79. Podemos notar uma diferença significativa nos valores do F1-score para a classe das secções intervencionadas, que comparativamente à mesma classe, mas das faixas ao longo de estradas, obteve menos 9%.

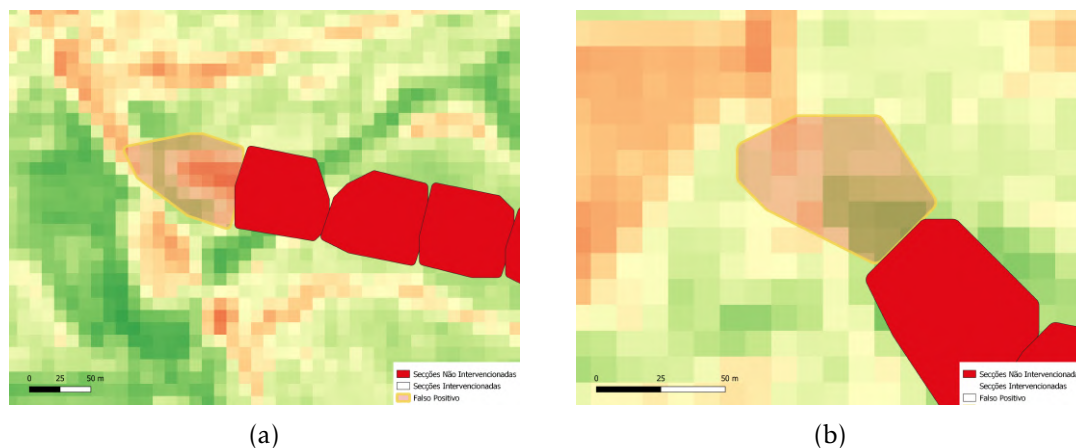


Figura 6.16: Exemplos de falsos dos positivos registados com o XGBoost e conjunto de dados *CD\_BANDAS*.

Dois exemplos de falsos de positivos obtidos com a classificação realizada pelo XGBoost e o conjunto de dados *CD\_BANDAS* podem ser visualizados na 6.16. Nestas duas secções onde não houve intervenção e foram classificadas como intervencionadas, podemos observar que parte das secções intersectam zonas onde não existe vegetação ou esta encontra-se seca, de tal maneira que esta pode ser a razão destas terem sido incorretamente classificadas.

### Sentinel-1

Foi analisado o desempenho dos dados do Sentinel-1 para as faixas ao longo de linhas elétricas. Foram testados os vários algoritmos usando as métricas mencionadas na secção 4.6 extraídas dos padrões temporais das polarizações VV e VH, e a diferença e o rácio entre elas.

Na Figura 6.17 estão as métricas F1-score e Kappa para todos os algoritmos testados. Para estas faixas os resultados foram bastante melhores que para as faixas das estradas usando os mesmos dados. O melhor resultado foi obtido pelo algoritmo RF que obteve um F1-score global de 0,79 e um Kappa de 0,57, sendo que o pior resultado foi o do algoritmo SVM. Embora os valores das métricas sejam mais baixas em comparação às obtidas com os dados do Sentinel-2, ainda assim foram boas, sendo o valor do F1-score sensivelmente 7% inferior.

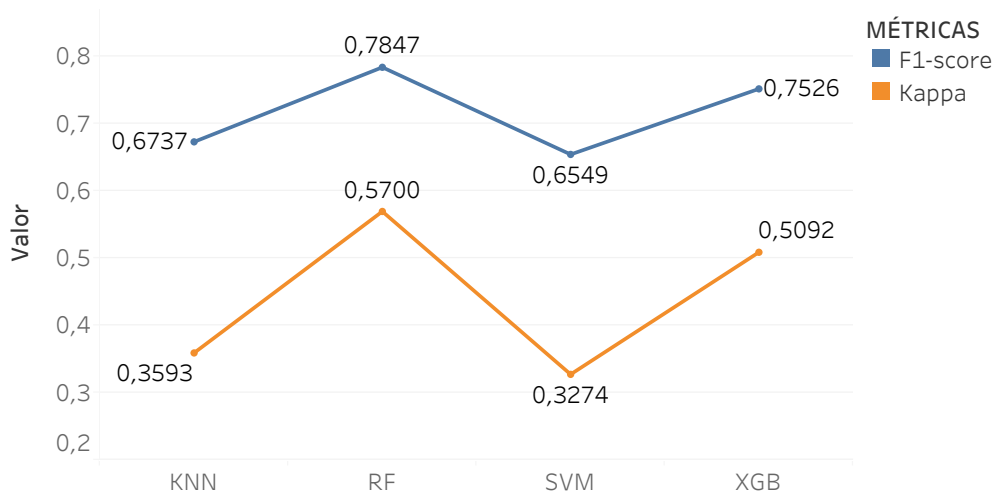


Figura 6.17: Valores das métricas F1-score e Kappa para os diferentes algoritmos usando os dados do Sentinel-1.

A matriz de confusão dos resultados obtidos pelo RF pode ser visualizada na Tabela 6.13. Tal como em todas as experiências anteriores a classe com os melhores resultados é a das secções não intervencionadas, com um F1-score de 0,90. A classe das secções intervencionadas obteve um F1-score de 0,67. Ao contrário do que foi observado usando o mesmo conjunto de dados nas faixas ao longo das estradas, o F1-score global não foi tão influenciado pelas secções não intervencionadas uma vez que as secções intervencionadas obtiveram um F1-score maior comparativamente ao das estradas.

Tabela 6.13: Matriz de confusão para o algoritmo RF usando os dados das polarizações do Sentinel-1.

		Referência		Total	Precisão	Recall	F1-score
		Interven.	Não Interven.				
Previsto	Interven.	114	45	159	0,72	0,62	0,67
	Não Interven.	69	529	598	0,89	0,92	0,90
	Total	183	574				
	F1-Score Global Kappa Global	0,79 0,57					

### Ambos os Satélites

Ao juntar os dados do Sentinel-1 com os dados do Sentinel-2 houve uma variação dos resultados em relação aos que foram obtidos usando apenas dados do Sentinel-2. Em alguns casos houve uma melhoria, enquanto que em outros os resultados desceram ligeiramente (Figura 6.18).

No RF e XGBoost os conjuntos de dados que obtiveram os melhores resultados com dados do Sentinel-2 desceram ligeiramente, sendo que para o XGBoost o *CD\_BANDAS* deixou de ser o conjunto com melhores resultados e passou a ser o *CD\_TUDO*. O maior impacto foi no conjunto de dados *CD\_ÍNDICES* com o algoritmo SVM que melhorou cerca de 5%. No algoritmo KNN, nenhum dos conjuntos foi muito afetado com a junção destes dois tipos de dados.

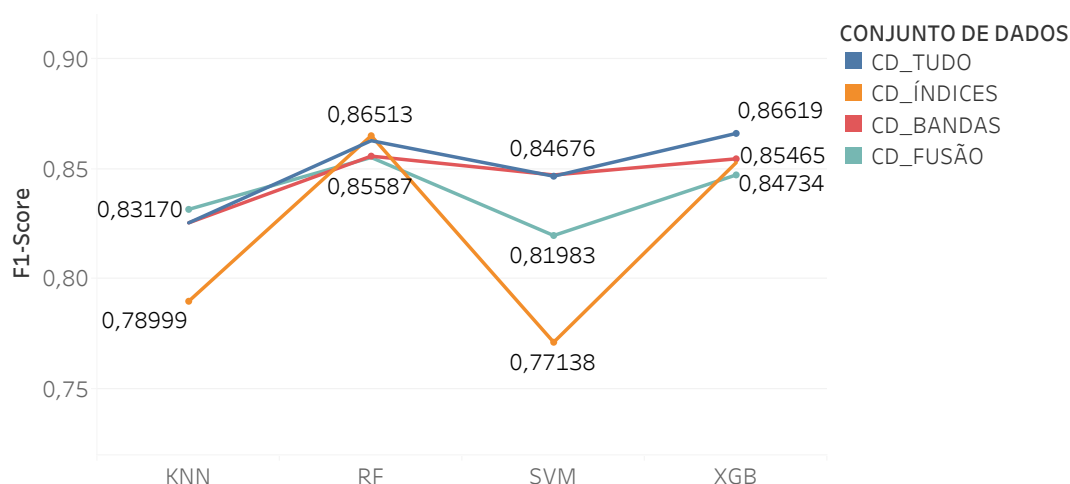


Figura 6.18: Valor da métrica F1-score para os diferentes algoritmos e conjuntos de dados com dados do Sentinel-1 e Sentinel-2.

A matriz de confusão para o algoritmo XGBoost com o conjunto *CD\_TUDO* pode ser consultada na Tabela 6.14. Como os resultados foram semelhantes aos observados anteriormente na experiência com os dados do Sentinel-2, o F1-score de ambas as classes é também igual, apesar dos valores da precisão e *recall* terem sido diferentes dos observados anteriormente.

Tabela 6.14: Matriz de confusão para o algoritmo XGBoost usando o conjunto de dados *CD\_TUDO*.

		Referência					
		Interven.	Não Interven.	Total	Precisão	Recall	F1-score
Previsto	Interven.	136	24	160	0,85	0,74	0,79
	Não Interven.	47	550	597	0,92	0,96	0,94
	Total	183	574				
	F1-Score Global Kappa Global	0,87 0,73					

No geral o impacto da junção dos dados do Sentinel-1 com os do Sentinel-2 não mostrou melhorias nos resultados dos algoritmos.

## 6.7 Importância das Características

Alguns conjuntos de dados usados têm um elevado número de características, que em alguns casos podem chegar a ser 340 (Tabela 6.2). Nos casos em que os conjuntos de dados têm um número de amostras bastante grande, o número de características pode ter um grande impacto nos tempos de processamento.

Foi usado o LASSO para tentar reduzir a dimensionalidade dos dados e ao mesmo tempo perceber o impacto que esta redução tem nos resultados dos algoritmos. Para tal considerou-se apenas o conjunto de dados *CD\_TUDO*, que gerou bons resultados na análise temporal e na estática, e foi aplicado o LASSO para os dados das faixas ao longo de estradas e ao longo de linhas elétricas.

O conjunto de dados *CD\_TUDO* na análise temporal e usando apenas os dados do Sentinel-2 tem um total de 260 características independentemente do tipo de faixa. Ao aplicar o LASSO o número de características desceu para menos de metade no entanto os resultados em alguns casos não foram muito afetados.

Nesta secção é analisado o impacto que a seleção de características teve na classificação e nos tempos de afinação de parâmetros dos diferentes algoritmos para as faixas ao longo de estradas e linhas elétricas.

### 6.7.1 Faixas ao Longo de Estradas

As faixas ao longo das estradas, para o conjunto de dados *CD\_TUDO* e usando dados temporais do Sentinel-2, tem um total de 260 características. Ao executar o LASSO para estes dados, as características foram reduzidas para 82. Na Figura 6.19 pode ser consultado o número de características selecionadas e eliminadas pelo LASSO, tendo em conta a banda ou o índice a partir do qual a característica foi calculada.

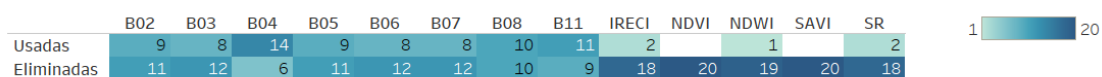


Figura 6.19: Número de características selecionadas e eliminadas pelo LASSO, agrupadas pelo tipo de dados (Bandas e índices de vegetação) obtidos do Sentinel-2.

Podemos observar que nenhuma das características calculadas a partir dos índices de vegetação NDVI e SAVI foi selecionada, e dos restantes índices foi usado um número muito baixo. As características calculadas a partir das bandas do satélite foram as mais usadas, com destaque para a banda 4 (banda no vermelho) que foi a que teve um maior número de características selecionadas.

Também foi feita a mesma análise mas para o tipo de característica (i.e. se a característica diz respeito ao interior da FGCI, ao *buffer* exterior, à diferença ( $dif_x$ ) ou rácio ( $ratio_x$ ) entre os dois, ou ao  $\Delta dif_k$ ) que pode ser consultada na Figura 6.20. Apenas foi selecionada uma característica que recorreu  $ratio_x$ , sendo que as características mais usadas dizem respeito a informação de dentro da faixa.

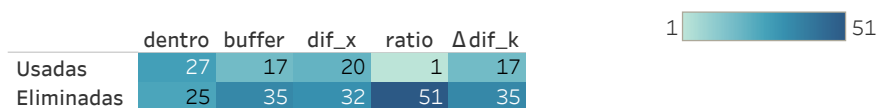


Figura 6.20: Número de características selecionadas e eliminadas pelo LASSO, agrupadas pelo tipo de característica.

Para avaliar o impacto que esta seleção tem nos resultados, foram executados os algoritmos usados com as 82 características selecionadas, e comparados com os resultados obtidos usando todas as 260 características. Na Figura 6.21 está presente o F1-score para o conjunto original de 260 características (azul) e para as características selecionadas (laranja).

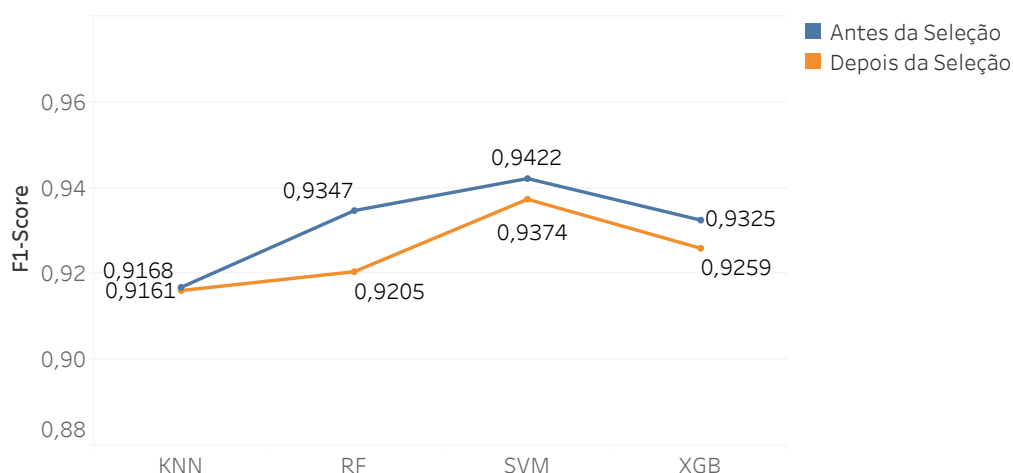


Figura 6.21: F1-score antes e depois de ter sido realizada a seleção de características.

Os resultados do algoritmo KNN não foram afetados pela redução do número de características, no entanto para todos os outros algoritmos houve uma redução do F1-score. A redução máxima observada foi de 1,5% no RF e a mínima no SVM com uma descida de apenas 0,5%. Contudo, tendo em conta que o número de características foi reduzido para 31% do total das características iniciais, os resultados mantiveram-se bastante aceitáveis.

Também foram realizados testes para analisar o impacto que a seleção de características teve nos tempos de afinação dos parâmetros dos algoritmos. Na Figura 6.22 estão presentes os tempos antes (azul) e depois (laranja) de ter sido realizada a seleção. O único algoritmo que não apresentou melhorias foi o XGBoost, para os restantes houve uma redução dos tempos para menos de metade.

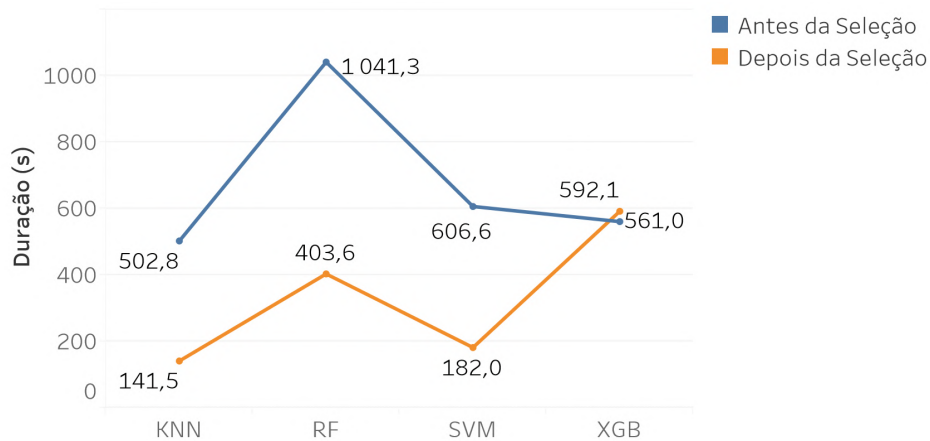


Figura 6.22: Tempos de afinação dos parâmetros dos algoritmos, antes e depois da seleção de características.

### 6.7.2 Faixas ao Longo de Linhas Elétricas

Tal como nas faixas ao longo das estradas, nas faixas ao longo das linhas elétricas também foi usado o conjunto de dados *CD\_TUDO* e como tal também tem um total de 260 características. Após executar o LASSO houve uma redução para 79 características. Na Figura 6.23 podemos ver o número de características selecionadas e eliminadas pelo LASSO, tendo em conta a banda ou o índice a partir do qual a característica foi calculada.

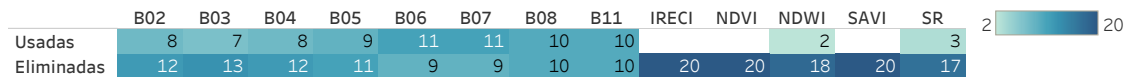


Figura 6.23: Número de características selecionadas e eliminadas pelo LASSO, agrupadas pelo tipo de dados (Bandas e índices de vegetação) obtidos do Sentinel-2.

Tal como aconteceu para as faixas ao longo de estradas, alguns índices de vegetação (NDVI, IRECI e SAVI) não tiveram nenhuma das suas características selecionadas pelo LASSO, e os que tiveram (NDWI e SR) foi um número bastante baixo apenas 1 ou 2. As características das bandas foram as mais selecionadas, embora neste caso estas digam respeito a bandas no *red-edge* (o RE2 e o RE3).

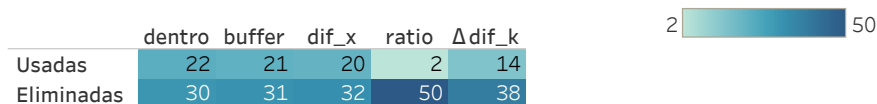


Figura 6.24: Número de características selecionadas e eliminadas pelo LASSO, agrupadas pelo tipo de característica.

Ao analisar número de características selecionadas, tendo em conta o seu tipo (Figura 6.24) podemos notar que mais uma vez o  $ratio_x$  foi o menos usado apenas com 2 características a serem selecionadas. As características com informação relativa ao interior da

FGCI, ao *buffer* e à *diff<sub>x</sub>* foram as mais selecionadas.

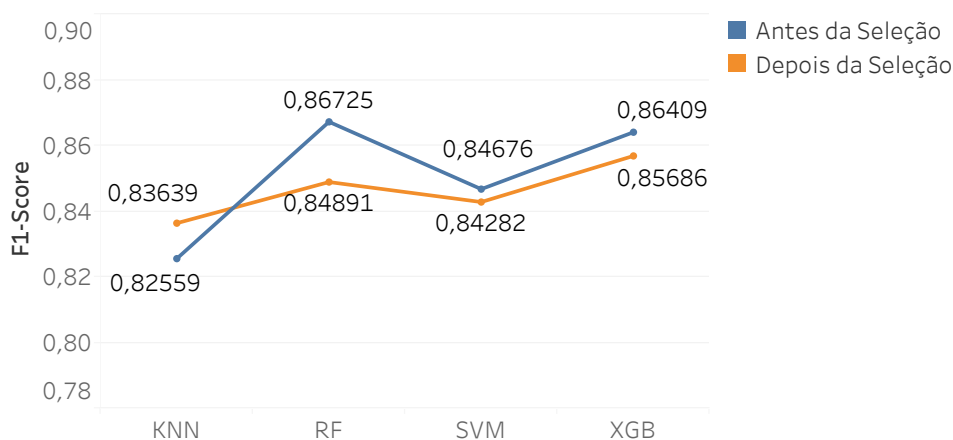


Figura 6.25: F1-score antes e depois de ter sido realizada a seleção de características.

De seguida foi analisado o impacto que esta seleção teve nos resultados. Foram executados os algoritmos com as 79 características selecionadas e comparados os resultados obtidos usando a totalidade das características. Os resultados podem ser observados na Figura 6.25. O KNN foi o único algoritmo em que os resultados melhoraram ao realizar a seleção de características, já os nos restantes algoritmos houve uma descida dos valores do F1-score que no caso do XGBoost e do RF foi inferior a 1%.

Os testes realizados para analisar o impacto que a seleção de características teve nos tempos de afinação dos parâmetros dos algoritmos podem ser visualizados na Figura 6.26. O comportamento observado nos tempos de afinação de parâmetros foi idêntico ao das faixas ao longo de estradas. Houve uma redução dos tempos de execução dos algoritmos para cerca de metade, com exceção do algoritmo XGBoost em que piorou ligeiramente.

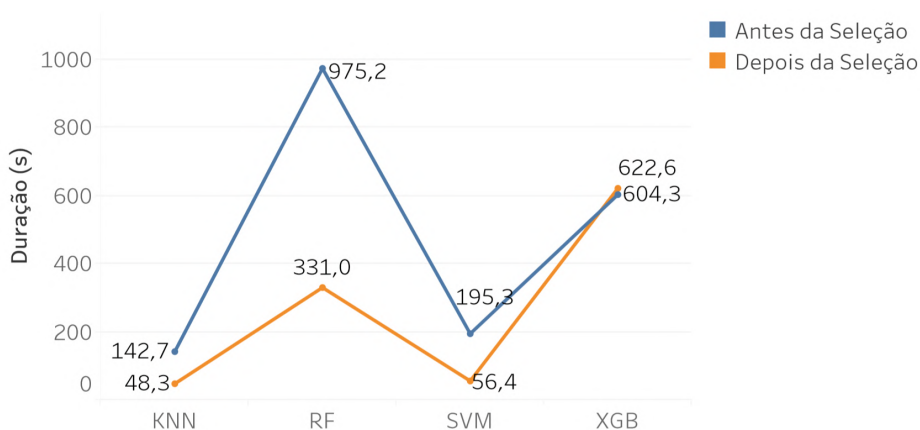


Figura 6.26: Tempos de afinação dos parâmetros dos algoritmos, antes e depois da seleção de características.



## 6.8 Estimativa de Datas de Intervenção

Foi realizada uma experiência com o objetivo de tentar estimar as datas em que ocorreram as intervenções nas FGCI. Foram usadas as faixas ao longo de estradas, e selecionadas algumas secções que foram intervencionadas na mesma altura do ano, mas em dois locais diferentes Local A (Figura 6.27a) e Local B (Figura 6.27b).



Figura 6.27: Secções usadas para estimar a data de intervenção e que foram intervencionadas na mesma altura do ano.

Foram escolhidas as faixas ao longo de estradas para realizar esta análise, pois a data exata da intervenção das faixas não estava disponível nos dados disponibilizados pela Câmara de Mação, e visualmente estas faixas são as que mais facilmente se nota o momento da intervenção através de imagens satélite. Podemos ver na Figura 6.28 o NDVI antes e depois de ter sido realizada a intervenção na FGCI no Local A e o mesmo para o Local B na Figura 6.29.

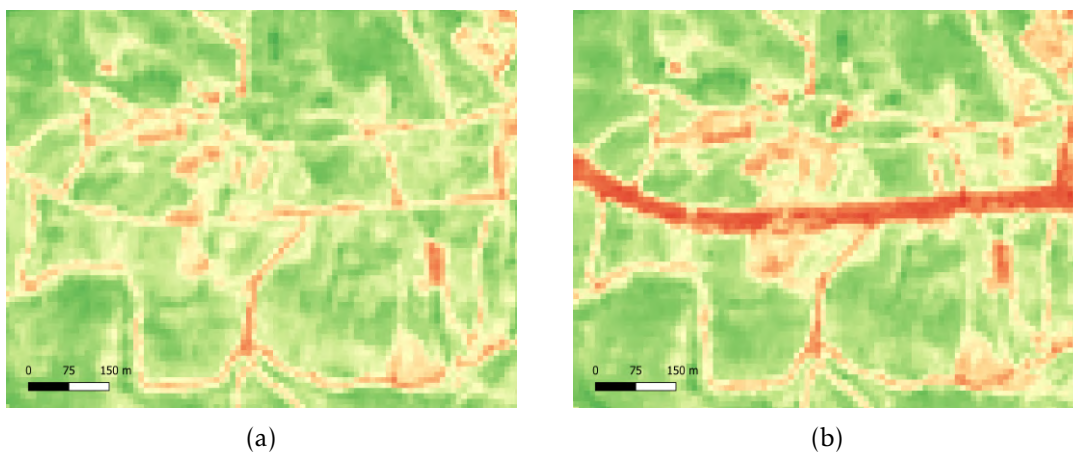


Figura 6.28: NDVI antes e depois da intervenção para o Local A. (a) NDVI dia 15-05-2018. (b) NDVI dia 19-06-2018.

Foi usado o padrão temporal da expressão  $\Delta dif_{NDVI}$  (Equação 4.3) que indica a diferença do  $dif_{NDVI}$  (Equação 4.1) entre duas datas consecutivas e estabelecendo um

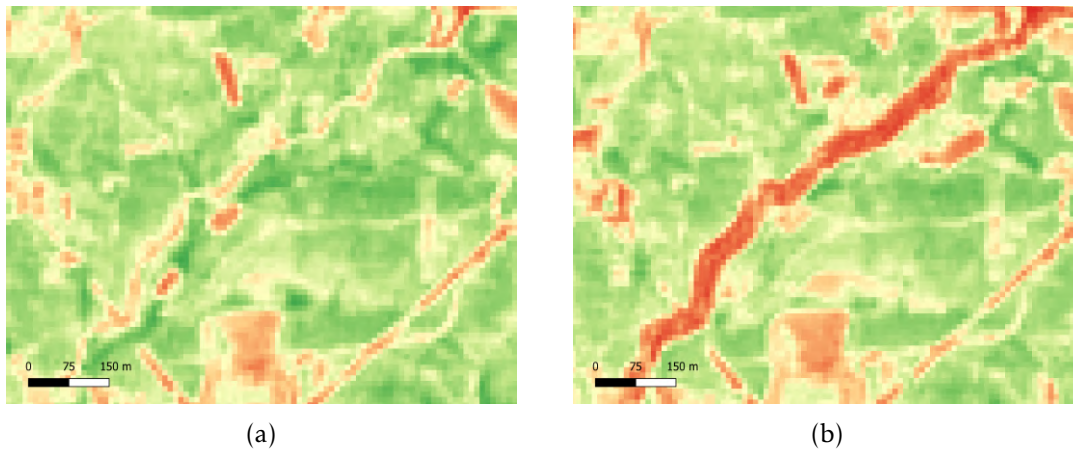


Figura 6.29: NDVI antes e depois da intervenção para o Local B. (a) NDVI dia 15-05-2018. (b) NDVI dia 19-06-2018.

limite  $l = 0,10$ , de tal forma que quando  $\Delta dif_{NDVI} > l$  é interpretado que a secção foi intervencionada. Este limite representa que houve uma descida dos valores do NDVI dentro da faixa de um mês para o outro, enquanto que no *buffer* exterior este índice manteve-se ou foi muito pouco afetado.

Na Figura 6.30 podemos ver o resultado desta abordagem para os dois locais analisados. Em ambos os casos a maioria das datas estimadas como datas de intervenção corresponderam com os dados observados, sendo que apenas 2 secções em 28 é que a estimativa foi incorreta, no entanto para várias secções não se obtiveram resultados.

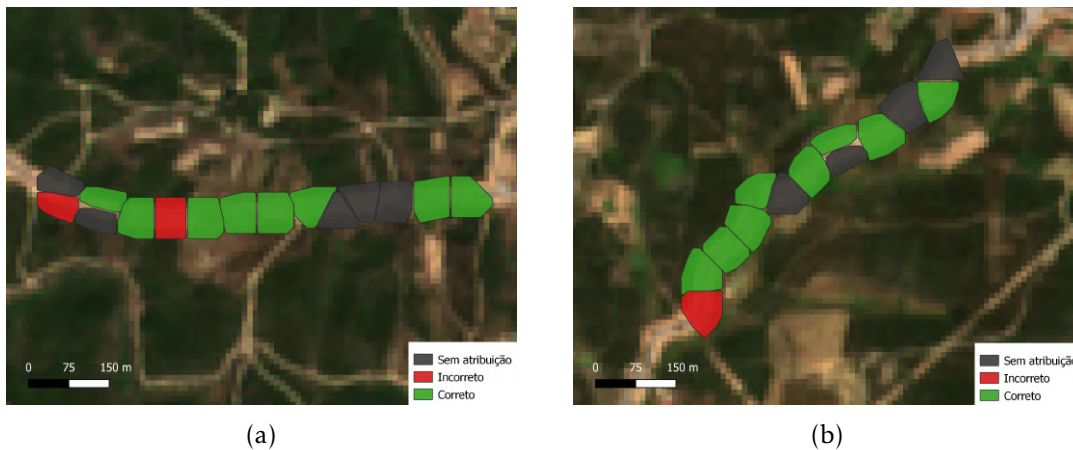


Figura 6.30: Resultados da previsão das datas para o Local A (a) e para o Local B (b). A vermelho secções onde a estimativa foi errada, a verde onde foi correta, e a preto secções que não obtiveram resultados.

No total 17 secções foram estimadas corretamente, 2 incorretamente e 9 secções não obtiveram uma atribuição da estimativa da data de intervenção. Este método revelou poder ser promissor, porém é necessário realizar mais experiências e ser feita uma análise mais detalhada de forma a obter uma abordagem mais robusta.

## 6.9 Discussão dos Resultados

Os resultados obtidos nas experiências realizadas de modo geral foram bastante bons. Realizou-se dois tipos de análise às FGCI: uma análise estática, recorrendo apenas a dados do Sentinel-2 de uma data, e uma análise temporal usando os padrões temporais das características do Sentinel-2 e do Sentinel-1.

Foram analisadas faixas ao longo de linhas elétricas e ao longo de estradas, propondo-se diferentes abordagens para classificar as secções destas faixas, desde o uso do algoritmo de agrupamento *K-Means* ao uso de algoritmos de classificação ou à classificação recorrendo ao comportamento dos padrões temporais, impondo um limite de forma a obter uma classificação.

Realizou-se também uma avaliação dos diferentes algoritmos de classificação, tanto na análise temporal como na análise estática, comparando os resultados destes algoritmos usando diferentes conjuntos de dados. Também se testou o algoritmo LASSO com a finalidade de tentar reduzir a dimensionalidade dos dados e analisar o impacto dessa redução.

Além destas experiências relacionadas com a classificação de faixas, foi também realizada uma experiência para tentar determinar as datas de intervenção das faixas ao longo de estradas. Nesta secção são discutidos os resultados das várias experiências realizadas.

### 6.9.1 Classificação das Intervenções nas FGCI

Realizaram-se dois tipos de análise para classificar as intervenções nas FGCI: a estática e a temporal. Em ambas as análises foi efetuada uma comparação das performances dos diferentes algoritmos de classificação usando diferentes conjuntos de dados, mas também considerou-se outras estratégias para detetar intervenções: agrupamento de dados e classificação através de imposição de limite nas séries temporais.

Com esta análise feita às FGCI foi possível avaliar vários fatores que contribuíram para a qualidade dos modelos gerados ou que influenciaram os resultados obtidos. Nesta secção são abordados alguns desses fatores e comparados os resultados obtidos.

#### 6.9.1.1 Algoritmos de Classificação

Foram utilizados diferentes algoritmos de classificação neste trabalho: o RF, SVM, XGBoost e KNN. Os algoritmos que se destacaram foram o XGBoost e o RF que em quase todas as experiências tiveram bons resultados. Porém o algoritmo SVM na análise temporal das faixas ao longo das estradas destacou-se obtendo o F1-score mais elevado, 0,94.

Os algoritmos que no geral obtiveram os piores resultados foram o SVM e o KNN. Na análise estática o SVM obteve os resultados mais baixos para ambos os tipos de faixa (linhas elétricas e estradas) usando o conjunto de dados *CD Índices*. No entanto, na análise temporal das faixas ao longo das estradas os resultados destes algoritmos melhoraram bastante, de tal forma que o SVM obteve o melhor resultado na experiência com

dados do Sentinel-2, mas com o XGBoost imediatamente atrás.

Na análise das faixas recorrendo aos dados temporais do Sentinel-1 o XGBoost foi o melhor a classificar as intervenções nas estradas (F1-score global de 0,79) enquanto que nas faixas das linhas elétricas o melhor foi o RF (F1-score de 0,85).

### 6.9.1.2 Métodos de Classificação Alternativos

Na análise estática foi realizado um agrupamento dos dados das secções usando o algoritmo *K-Means*. Foi usada uma zona no norte do concelho de Mação para realizar um agrupamento de secções de faixas ao longo de estradas e ao redor de habitações. Ao longo das estradas foi possível comparar os resultados deste agrupamento com os dados oficiais das intervenções, o que mostrou que o agrupamento gerado pelo *K-Means* esta correlacionado com os dados oficiais de referência, gerando um F1-score de 0,80 nas secções intervencionadas e de 0,75 nas não intervencionadas. Nas faixas ao redor das habitações não existiam dados de referência para comparação, ainda assim foi possível observar a existência de uma relação entre o agrupamento gerado e zonas onde não existe vegetação e zonas onde existe vegetação.

Na análise temporal, além do uso de algoritmos de classificação, foi usado um método que através das séries temporais de índices de vegetação e usando o  $dif_x$  classifica uma secção como intervencionada se o  $dif_x$  for superior a um dado valor de limite. Apesar de ser um método bastante simples os resultados obtidos foram razoáveis, chegando a obter um F1-score de 0,69 para as secções intervencionadas usando a série temporal do  $dif_{NDVI}$ .

Estes dois métodos usados divergem das abordagens que costumam ser empregues para este tipo de aplicação, porém geraram resultados interessantes e foram úteis para perceber melhor as especificidades dos dados que estavam a ser usados.

### 6.9.1.3 Impacto dos Conjuntos de Dados

Foram comparados 4 conjuntos de dados diferentes: *CD\_TUDO*, *CD\_ÍNDICES*, *CD\_BANDAS*, *CD\_FUSÃO*. Na análise estática foram apenas usados dados do Sentinel-2 e na análise temporal foi adicionada informação do Sentinel-1.

Tanto na análise temporal como na análise estática o conjunto de dados *CD\_TUDO* esteve sempre entre os conjuntos de dados a obter as métrica mais elevadas, sendo mesmo o melhor em boa parte das experiências.

O conjunto de dados *CD\_ÍNDICES* foi o que gerou os piores resultados. Na análise estática e para quase todos os algoritmos, este foi o conjunto que gerou os resultados mais baixos. Já na análise temporal, para os algoritmos RF e XGBoost o *CD\_ÍNDICES* obteve um F1-score próximo dos melhores resultados destes algoritmos, no entanto para o SVM e KNN os resultados foram bastante baixos, chegando a haver uma diferença de 23% para o melhor resultado num dos testes (usando o KNN nas faixas ao longo das estradas).

Como o *CD\_ÍNDICES* obteve em quase todas as experiências piores resultados que o conjunto de dados *CD\_BANDAS*, muitas vezes com uma grande diferença entre os resultados de ambos, podemos concluir que os dados dos índices de vegetação não tiveram uma contribuição importante para melhorar os resultados dos classificadores.

Os resultados das classificações usando apenas dados do Sentinel-1 foram mistos. No caso das estradas quase nenhuma secção intervencionada foi classificada corretamente pelo KNN e no caso do SVM não houve mesmo nenhuma secção classificada como intervencionada. Nas linhas elétricas os resultados foram melhores nas secções intervencionadas (F1-score de 0,67). Em ambas as faixas as secções não intervencionadas obtiveram resultados elevados.

A combinação dos dados do Sentinel-1 com os do Sentinel-2 não mostrou grandes melhorias nos resultados, apenas em alguns casos em que os resultados eram bastante baixos antes da junção é que houve uma melhoria significativa das métricas (aumentando F1-score em 4% ).

#### 6.9.1.4 Faixas Analisadas

Foram analisadas as faixas ao longo de estradas e linhas elétricas no concelho de Mação e foram usados algoritmos de classificação para tentar prever quais as secção destas faixas que foram intervencionadas.

Os resultados obtidos foram diferentes para ambos os tipos de faixa. Nas faixas ao longo de estradas, e usando séries temporais dos dados do Sentinel-2, foi obtido um F1-score global de 0,94, no entanto para as faixas ao longo de linhas elétricas o valor foi de 0,87. Ao analisar mais detalhadamente podemos notar que a classe das secções intervencionadas foi onde houve a maior diferença ( cerca de 10% ) entre estes dois tipos de faixa, enquanto que na classe das secções não intervencionadas a diferença foi de apenas de 5%.

Uma das razões que podem ser a causa desta diferença de valores é o facto das faixas ao longo das linhas elétricas serem mantidas ao longo do ano, já há vários anos, devido ao risco de ignição de incêndios que está associado a estas faixas. Como a vegetação dentro deste tipo de faixa já é reduzida antes da intervenção, quando é realizado algum tipo de intervenção a mudança não é tão drástica, e não será muito notável no sinal do  $dif_x$  4.1. Além disto, o facto de uma secção deste tipo de faixas conter menos informação que uma secção de faixas ao longo de estradas, pode ter algum impacto fazendo com que os resultados destas faixas sejam inferiores.

Por outro lado, nas faixas ao longo das estradas não existe a necessidade de manter a vegetação ao longo do ano, e as faixas deste tipo que foram analisadas aparentavam ter alguma quantidade de vegetação antes da intervenção, fazendo com que a intervenção fosse bem notável através do sinal do  $dif_x$ .

No geral as classificações geradas para estas faixas foram boas e mostraram que o uso de técnicas de deteção remota em junção com algoritmos de classificação podem ajudar

da detecção de faixas intervencionadas.

### 6.9.2 Seleção de características

A seleção de características realizada usando o LASSO gerou resultados interessantes, uma vez que alguns resultados não foram afetados, mesmo quando o número de dimensões do conjunto de dados foi reduzido para menos de metade. Além disso houve uma clara melhoria dos tempos de afinação de parâmetros para a maioria dos algoritmos, com a exceção do XGBoost em que os tempos pioraram ligeiramente.

No que diz respeito às características que foram selecionadas usando este método, os resultados mostram que as características calculadas a partir dos índices de vegetação não foram considerados relevantes, sendo que em alguns casos nenhuma característica dos índices foi selecionada. Por outro lado, as características das bandas do Sentinel-2 foram as mais relevantes, sendo que no caso das estradas a que se destacou foi a banda no vermelho (B04) e nas faixas ao longo das linhas elétricas as bandas no *red-edge*. Estes resultados vão de acordo com o que foi observado na comparação de conjunto de dados, onde os índices de vegetação não mostraram contribuir para a melhoria das classificações.

Em relação ao tipo de informação extraído das características ( $dif_x$ ,  $ratio_x$ ,  $\Delta dif_k$ ,  $buffer$  e dentro da FGCI) o  $ratio_x$  foi o menos selecionado tanto nas faixas ao longo das estradas como nas faixas ao longo das linhas. Nas faixas das estradas as características com informação referente ao interior da faixa foram as mais selecionadas e nas faixas ao longo das linhas o número de características selecionadas foi semelhante para o  $buffer$ , o  $dif_x$  e para dados de dentro da faixa.

Estes resultados mostram que pode ser benéfico o uso deste tipo de técnica, especialmente em conjuntos de dados que têm uma grande dimensionalidade, para perceber o impacto das características e reduzir os tempos associados com a afinação de parâmetros dos algoritmos.

### 6.9.3 Datas de Intervenção

Foi realizada uma experiência que visou estimar as datas de intervenção de algumas secções ao longo das faixas das estradas.

Foram selecionados dois locais onde as faixas foram intervencionadas na mesma altura do ano e foram marcadas algumas secções. Usando o comportamento da série temporal do  $\Delta dif_k$  4.3 e estabelecendo um limite foi possível realizar algumas estimativas sobre as datas de intervenção.

Apesar do número de secções analisadas nesta experiência ser reduzido (28 secções), este método acertou corretamente em 17 delas, sendo que apenas 2 foram incorretamente estimadas. Porém, 9 secções não obtiveram uma atribuição de data o que indica que o limite usado pode não ser o mais indicado.

Apesar da experiência ter gerado resultados promissores, a amostra de secções foi pequena e o método pode ainda ser melhorado, sendo necessário realizar mais testes e fazer uma análise mais profunda.





## CONCLUSÃO E TRABALHO FUTURO

### 7.1 Conclusão

Nesta dissertação foi abordado o tema das FGCI, mais em específico da detecção de intervenções dentro destas faixas. Estas faixas que têm como função proteger infraestruturas, dificultar a progressão do fogo ou prevenir a ignição de incêndios, necessitam de estar devidamente limpas de acordo com as regras estipuladas. Devido às mudanças da vegetação ao longo do ano e à elevada extensão das redes de FGCI, a tarefa de monitorização das faixas intervencionadas pode ser difícil.

Foi apresentado um método para detecção de intervenções nas FGCI que recorre a algoritmos de aprendizagem automática e a imagens de satélite do Sentinel-2 e Sentinel-1 para determinar se uma faixa foi ou não intervencionada.

A abordagem utilizada para detetar uma intervenção consiste na comparação de vegetação entre o interior e exterior de uma faixa de forma a detetar mudanças que ocorram na vegetação do interior das FGCI. Esta abordagem exige a criação de *buffers* exteriores às faixas, em zonas de vegetação, que dependendo do tipo de faixas podem ser complexos de criar.

Adotaram-se duas abordagens de análise, uma com o uso de apenas uma imagem para realizar as classificações e outra recorrendo ao uso de séries temporais. Apesar da análise estática ter mostrado obter bons resultados, principalmente para as faixas ao longo de estradas (F1-score 0,91), o uso de séries temporais melhorou os resultados fazendo subir o F1-score, no caso das faixas ao longo das estradas, de 0,91 para 0,94.

Foram testados os algoritmos KNN, RF, SVM e XGBoost. Os algoritmos que se destacaram foram o RF e o XGBoost, com os seus resultados a serem os melhores em praticamente todos os testes realizados, com a exceção na análise temporal das faixas ao longo das estradas em que o SVM foi o melhor, mas com valores semelhantes ao RF e XGBoost.

Os conjuntos de dados comparados obtiveram diferentes resultados. O conjunto que no geral gerou as melhores classificações para os vários algoritmos foi o *CD\_TUDO* sendo que os piores resultados foram gerados usando o conjunto de dados *CD\_ÍNDICES*, indicando que os dados dos índices de vegetação não contribuíram para a melhorar as classificações. O mesmo foi observado na seleção de características onde as características criadas a partir dos índices de vegetação não foram selecionadas pelo LASSO.

Os melhores resultados para classificação das faixas intervencionadas foram obtidos com as séries temporais dos dados do Sentinel-2. Para as faixas ao longo das estradas foi obtido um F1-score global de 0,94 e um Kappa global de 0,88 usando o algoritmos SVM com o conjunto de dados *CD\_TUDO*. Nas faixas ao longo de linhas elétricas os resultados foram mais baixos com o F1-score global de 0,87 e um kappa global de 0,74.

A discrepância dos resultados entre estas duas faixas pode ser justificada pela diferença nas intervenções realizadas nestas faixas. A manutenção das faixas ao longo de linhas elétricas é realizada há vários anos, sendo que a vegetação necessita de ser controlada ao longo do ano devido ao risco de descargas elétricas. Nas faixas ao longo das estradas não existe este tipo de necessidade, o que potencia o crescimento de vegetação ao longo do ano. Assim, as mudanças na vegetação após uma intervenção são mais facilmente distinguíveis em faixas ao longo de estradas do que em faixas ao longo de linhas elétricas.

Resultou deste trabalho uma abordagem para detetar intervenções nas FGCI, sendo que os modelos criados conseguiram estimar com uma boa precisão as faixas que foram intervencionadas na zona analisada. Este método aqui apresentado pode ser aplicado em outras zonas do país, embora esteja dependente da qualidade da informação vetorial das FGCI.

## 7.2 Trabalho Futuro

Apesar dos resultados obtidos neste trabalho terem sido bastante bons na sua globalidade, existem alguns aspetos que podem ser melhorados, e informação que pode ser adicionada para melhorar alguns dos modelos.

Informação sobre a elevação do terreno pode ser pertinente. Para as faixas ao longo de linhas elétricas, existem situações em que a vegetação não necessita de ser removida por baixo das linhas, essas situações ocorrem em zonas de vales acentuados em que as linhas elétricas passam a várias dezenas de metros do solo. Estas situações podem induzir em erro na análise dos modelos uma vez que as secções podem estar indicadas como intervencionadas e ser uma destas situações. Assim ao introduzir informação como o relevo ou declive do terreno pode ajudar a melhorar os resultados dos modelos para o caso específico das faixas ao longo de linhas elétricas.

Outra situação que pode afetar os resultados, especialmente quando estão a ser usadas séries temporais, é as sombras que podem ocorrer em diferentes alturas do ano, especialmente no inverno quando o ângulo de incidência do sol é mais baixo. Estas sombras

podem ser causadas por infraestruturas artificiais ou naturais como montanhas ou até árvores de grandes dimensões. Nestes casos os valores dos índices de vegetação são afetados e podem induzir em erro. Assim, informação sobre estas áreas que durante o período de inverno ficam na sombra pode ajudar a melhorar os modelos.

A partir dos dados do satélite Sentinel-1 podem ser calculadas outras características, como por exemplo a textura, que neste trabalho não foram utilizadas e podem ter impacto nos resultados obtidos com os dados deste satélite.



## BIBLIOGRAFIA

- [1] S. Abdikan, F. B. Sanli, M. Ustuner e F. Calo. “LAND COVER MAPPING USING SENTINEL-1 SAR DATA”. Em: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B7.July (jun. de 2016), pp. 757–761. DOI: [10.5194/isprsarchives-XLI-B7-757-2016](https://doi.org/10.5194/isprsarchives-XLI-B7-757-2016). URL: <http://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLI-B7/757/2016/isprs-archives-XLI-B7-757-2016.pdf>.
- [2] A. Agapiou, D. G. Hadjimitsis, C. Papoutsas, D. D. Alexakis e G. Papadavid. “The Importance of accounting for atmospheric effects in the application of NDVI and interpretation of satellite imagery supporting archaeological research: The case studies of Palaepaphos and Nea Paphos sites in Cyprus”. Em: *Remote Sensing* 3.12 (2011), pp. 2605–2629. ISSN: 20724292. DOI: [10.3390/rs3122605](https://doi.org/10.3390/rs3122605).
- [3] M. da Agricultura do Desenvolvimento Rural e das Pescas. *Decreto-Lei n.º 124/2006*. Diário da República n.º 123/2006, Série I-A de 2006-06-28. 2006. URL: <https://data.dre.pt/eli/dec-lei/124/2006/06/28/p/dre/pt/html>.
- [4] *Área ardida em Portugal (1980-2017)*. URL: <https://www.pordata.pt/Portugal/Inc%C3%AAndios+rurais+e+%C3%A1rea+ardida+%E2%80%5C93+Continente-1192> (acedido em 09/09/2019).
- [5] N. Baghdadi, M. E. Hajj e M. Zribi. “Coupling SAR C-band and optical data for soil moisture and leaf area index retrieval over irrigated grasslands”. Em: *International Geoscience and Remote Sensing Symposium (IGARSS)*. Vol. 2016-Novem. 3. 2016, pp. 3551–3554. ISBN: 9781509033324. DOI: [10.1109/IGARSS.2016.7729919](https://doi.org/10.1109/IGARSS.2016.7729919).
- [6] M. Belgiu e O. Csillik. “Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis”. Em: *Remote Sensing of Environment* 204.January 2017 (jan. de 2018), pp. 509–523. ISSN: 00344257. DOI: [10.1016/j.rse.2017.10.005](https://doi.org/10.1016/j.rse.2017.10.005). URL: <https://doi.org/10.1016/j.rse.2017.10.005>  
<https://linkinghub.elsevier.com/retrieve/pii/S0034425717304686>.
- [7] J. A. A. Castillo, A. A. Apan, T. N. Maraseni e S. G. Salmo. “Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery”. Em: *ISPRS Journal of Photogrammetry and Remote Sensing* 134 (dez. de 2017), pp. 70–85. ISSN: 0924-2716. DOI: [10.1016/j.isprsjprs.2017.12.005](https://doi.org/10.1016/j.isprsjprs.2017.12.005).

- J. ISPRS JPRS. 2017. 10. 016. URL: <https://www.sciencedirect.com/science/article/pii/S0924271617303362?via%7B%5C%7D3Dihub>.
- [8] T. Chen e C. Guestrin. "XGBoost: A Scalable Tree Boosting System". Em: *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 785–794. DOI: 10.1145/2939672.2939785. URL: <http://doi.acm.org/10.1145/2939672.2939785>.
- [9] J. G. Clevers, L. Kooistra e M. M. van den Brande. "Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop". Em: *Remote Sensing* 9.5 (2017), pp. 1–15. ISSN: 20724292. DOI: 10.3390/rs9050405.
- [10] I. da Conservação da Natureza e das Florestas - ICNF. *Manual de Rede Primária*. 2014. URL: <http://www2.icnf.pt/portal/florestas/dfci/Resource/doc/cartografia-dfci/manual-RPFGC-20mai2014.pdf> (acedido em 29/01/2019).
- [11] C. Dechoz, V. Poulain, S. Massera, F. Languille, D. Greslou, F. de Lussy, A. Gaudel, C. L'Helguen, C. Picard e T. Trémas. "Sentinel 2 global reference image". Em: *Image and Signal Processing for Remote Sensing XXI*. Vol. 9643. International Society for Optics e Photonics. 2015, 96430A.
- [12] J. J. Erinjery, M. Singh e R. Kent. "Mapping and assessment of vegetation types in the tropical rainforests of the Western Ghats using multispectral Sentinel-2 and SAR Sentinel-1 satellite imagery". Em: *Remote Sensing of Environment* 216 (out. de 2018), pp. 345–354. DOI: 10.1016/j.rse.2018.07.006. URL: <https://www.sciencedirect.com/science/article/pii/S003442571830333X>.
- [13] e. a. Evgenii A. Maltsev Yuri A. Maglinets. "Development of the Technology to Identify Firebreak Plowing Objects Based on the Satellite Data of the Earth Remote Sensing". Em: *Journal of Siberian Federal University*. (2018).
- [14] W. J. Frampton, J. Dash, G. Watmough e E. J. Milton. "Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation". Em: *ISPRS Journal of Photogrammetry and Remote Sensing* 82 (2013), pp. 83–92. DOI: 10.1016/j.isprsjprs.2013.04.007.
- [15] B. C. Gao. "NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space". Em: *Remote Sensing of Environment* 58.3 (dez. de 1996), pp. 257–266. DOI: 10.1016/S0034-4257(96)00067-3. URL: <https://www.sciencedirect.com/science/article/pii/S0034425796000673>.
- [16] GDAL - biblioteca para transformações geoespaciais de vetores e de dados raster. URL: <https://www.gdal.org/> (acedido em 14/02/2019).
- [17] C. Gómez, J. C. White e M. A. Wulder. *Optical remotely sensed time series data for land cover classification: A review*. Jun. de 2016. DOI: 10.1016/j.isprsjprs.2016.03.008.

- [18] D. G. Hadjimitsis, G. Papadavid, A. Agapiou, K. Themistocleous, M. G. Hadjimitsis, A. Retalis, S. Michaelides, N. Chrysoulakis, L. Toullos e C. R. I. Clayton. “Atmospheric correction for satellite remotely sensed data intended for agricultural applications: Impact on vegetation indices”. Em: *Natural Hazards and Earth System Science* 10.1 (2010), pp. 89–95. ISSN: 16849981. DOI: [10.5194/nhess-10-89-2010](https://doi.org/10.5194/nhess-10-89-2010). arXiv: [arXiv:1409.3476v2](https://arxiv.org/abs/1409.3476v2).
- [19] S. S. Heydari e G. Mountrakis. “Effect of classifier selection, reference sample size, reference class distribution and scene heterogeneity in per-pixel classification accuracy using 26 Landsat sites”. Em: *Remote Sensing of Environment* 204. February 2017 (jan. de 2018), pp. 648–658. DOI: [10.1016/j.rse.2017.09.035](https://doi.org/10.1016/j.rse.2017.09.035). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0034425717304601>.
- [20] A. Huete. “A soil-adjusted vegetation index (SAVI)”. Em: *Remote Sensing of Environment* 25.3 (ago. de 1988), pp. 295–309. DOI: [10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X). URL: <https://www.sciencedirect.com/science/article/abs/pii/003442578890106X>.
- [21] M. Immitzer, F. Vuolo e C. Atzberger. “First experience with Sentinel-2 data for crop and tree species classifications in central Europe”. Em: *Remote Sensing* 8.3 (2016). DOI: [10.3390/rs8030166](https://doi.org/10.3390/rs8030166).
- [22] J. Inglada, M. Arias, B. Tardy, O. Hagolle, S. Valero, D. Morin, G. Dedieu, G. Sepulcre, S. Bontemps, P. Defourny e B. Koetz. “Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery”. Em: *Remote Sensing* 7.9 (2015), pp. 12356–12379. ISSN: 20724292. DOI: [10.3390/rs70912356](https://doi.org/10.3390/rs70912356).
- [23] A. Interna. *Decreto-Lei n.º 10/2018*. Diário da República n.º 32/2018, Série I de 2018-02-14. 2018. URL: <https://data.dre.pt/eli/dec-lei/10/2018/02/14/p/dre/pt/html>.
- [24] H. Jiang, D. Li, W. Jing, J. Xu, J. Huang, J. Yang e S. Chen. “Early Season Mapping of Sugarcane by Applying Machine Learning Algorithms to Sentinel-1A/2 Time Series Data: A Case Study in Zhanjiang City, China”. Em: *Remote Sensing* 11.7 (2019). ISSN: 2072-4292. DOI: [10.3390/rs11070861](https://doi.org/10.3390/rs11070861). URL: <https://www.mdpi.com/2072-4292/11/7/861>.
- [25] A. Moreira e M. Y. Santos. “Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points”. Em: *GRAPP*. 2007.
- [26] H. Müller, P. Rufin, P. Griffiths, A. J. Barros Siqueira e P. Hostert. “Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape”. Em: *Remote Sensing of Environment* 156 (jan. de 2015), pp. 490–499. ISSN: 00344257. DOI: [10.1016/j.rse.2014.10.014](https://doi.org/10.1016/j.rse.2014.10.014). URL: <https://www.sciencedirect.com/science/article/pii/S0034425714004246>.

- [27] P. T. Noi e M. Kappas. “Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery”. Em: *Sensors (Switzerland)* 18.1 (2018). ISSN: 14248220. DOI: [10.3390/s18010018](https://doi.org/10.3390/s18010018).
- [28] *OpenCV - Biblioteca de visão computacional e software de aprendizagem automática*. URL: <https://opencv.org/> (acedido em 28/05/2019).
- [29] C. Pelletier, S. Valero, J. Inglada, N. Champion e G. Dedieu. “Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas”. Em: *Remote Sensing of Environment* 187 (2016), pp. 156–168. ISSN: 00344257. DOI: [10.1016/j.rse.2016.10.010](https://doi.org/10.1016/j.rse.2016.10.010). URL: <http://dx.doi.org/10.1016/j.rse.2016.10.010>.
- [30] M. A. Peña, R. Liao e A. Brenning. “Using spectrottemporal indices to improve the fruit-tree crop classification accuracy”. Em: *ISPRS Journal of Photogrammetry and Remote Sensing* 128 (jun. de 2017), pp. 158–169. DOI: [10.1016/j.isprsjprs.2017.03.019](https://doi.org/10.1016/j.isprsjprs.2017.03.019). URL: <https://www.sciencedirect.com/science/article/pii/S0924271616305019>.
- [31] M. Piragnolo, G. Lusiani e F. Pirotti. “Comparison of vegetation indices from RPAS and Sentinel-2 imagery for detecting permanent pastures”. Em: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 42.3 (2018), pp. 1381–1387. DOI: [10.5194/isprs-archives-XLII-3-1381-2018](https://doi.org/10.5194/isprs-archives-XLII-3-1381-2018).
- [32] *PMDFGCI - Guia Técnico*. URL: <http://www2.icnf.pt/portal/florestas/dfci/Resource/doc/Guia-Tecnico-PMDFGCI-AFN-Abril2012-v1.pdf> (acedido em 15/08/2019).
- [33] Y. Qian, W. Zhou, J. Yan, W. Li e L. Han. “Comparing machine learning classifiers for object-based land cover classification using very high resolution imagery”. Em: *Remote Sensing* 7.1 (2015), pp. 153–168. DOI: [10.3390/rs70100153](https://doi.org/10.3390/rs70100153).
- [34] A. Romero, C. Gatta e G. Camps-Valls. “Unsupervised deep feature extraction for remote sensing image classification”. Em: *IEEE Transactions on Geoscience and Remote Sensing* 54.3 (2016), pp. 1349–1362. ISSN: 01962892. DOI: [10.1109/TGRS.2015.2478379](https://doi.org/10.1109/TGRS.2015.2478379). arXiv: [1511.08131](https://arxiv.org/abs/1511.08131).
- [35] O. Rozenstein, N. Haymann, G. Kaplan e J. Tanny. “Estimating cotton water consumption using a time series of Sentinel-2 imagery”. Em: *Agricultural Water Management* 207 (ago. de 2018), pp. 44–52. DOI: [10.1016/j.agwat.2018.05.017](https://doi.org/10.1016/j.agwat.2018.05.017). URL: <https://www.sciencedirect.com/science/article/pii/S0378377418304785>.
- [36] M. T. S Clerc. *Sentinel-2 L1C Data Quality Report*. 2019. URL: [https://sentinel.esa.int/documents/247904/685211/Sentinel-2\\_L1C\\_Data\\_Quality\\_Report](https://sentinel.esa.int/documents/247904/685211/Sentinel-2_L1C_Data_Quality_Report) (acedido em 20/09/2019).



- 
- [37] V. C. D. Sarantis-Angelos G. Liampas Christos C. Stamatiou. "Comparison of three DEM sources: a case study from Greek forests". Em: vol. 10773. 2018.
- [38] O. Satir e S. Berberoglu. "Crop yield prediction under soil salinity using satellite derived vegetation indices". Em: *Field Crops Research* 192 (jun. de 2016), pp. 134–143. DOI: [10.1016/j.fcr.2016.04.028](https://doi.org/10.1016/j.fcr.2016.04.028). URL: <https://www.sciencedirect.com/science/article/pii/S0378429016301204>.
- [39] *Sen2Cor- Processor for Sentinel-2 Level 2A product generation and formatting*. URL: <http://step.esa.int/main/third-party-plugins-2/sen2cor/> (acedido em 29/01/2019).
- [40] *Sentinel-1 data products*. URL: <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-1-sar/products-algorithms/level-1-algorithms/products> (acedido em 16/02/2019).
- [41] *Sentinel-2 product types*. URL: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/product-types> (acedido em 15/02/2019).
- [42] *Sentinel-2 Radiometric Resolutions*. URL: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/radiometric> (acedido em 15/02/2019).
- [43] T. D. Setiyono, E. D. Quicho, L. Gatti, M. CamposTaberner, L. Busetto, F. Collivignarelli, F. J. GarciaHaro, M. Boschetti, N. I. Khan e F. Holecz. "Spatial rice yield estimation based on MODIS and Sentinel-1 SAR data and ORYZA crop growth model". Em: *Remote Sensing* 10.2 (2018), pp. 1–20. DOI: [10.3390/rs10020293](https://doi.org/10.3390/rs10020293).
- [44] M. Sibanda, O. Mutanga e M. Rouget. "Examining the potential of Sentinel-2 MSI spectral resolution in quantifying above ground biomass across different fertilizer treatments". Em: *ISPRS Journal of Photogrammetry and Remote Sensing* (2015). ISSN: 09242716. DOI: [10.1016/j.isprsjprs.2015.10.005](https://doi.org/10.1016/j.isprsjprs.2015.10.005). eprint: [arXiv:1011.1669v3](https://arxiv.org/abs/1011.1669v3).
- [45] T. Takayama e A. Iwasaki. "OPTIMAL WAVELENGTH SELECTION on HYPERSPECTRAL DATA with FUSED LASSO for BIOMASS ESTIMATION of TROPICAL RAIN FOREST". Em: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 3.July (2016), pp. 101–108. ISSN: 21949050. DOI: [10.5194/isprs-annals-III-8-101-2016](https://doi.org/10.5194/isprs-annals-III-8-101-2016).
- [46] S. Vafaei, J. Soosani, K. Adeli, H. Fadaei, H. Naghavi, T. D. Pham e D. T. Bui. "Improving accuracy estimation of Forest Aboveground Biomass based on incorporation of ALOS-2 PALSAR-2 and Sentinel-2A imagery and machine learning: A case study of the Hyrcanian forest area (Iran)". Em: *Remote Sensing* 10.2 (2018). DOI: [10.1007/s11401-018-1063-7](https://doi.org/10.1007/s11401-018-1063-7).

- [47] L. Yan, D. P. Roy, Z. Li, H. K. Zhang e H. Huang. “Sentinel-2A multi-temporal misregistration characterization and an orbit-based sub-pixel registration methodology”. Em: *Remote Sensing of Environment* 215 (set. de 2018), pp. 495–506. DOI: [10.1016/j.rse.2018.04.021](https://doi.org/10.1016/j.rse.2018.04.021). URL: <https://www.sciencedirect.com/science/article/pii/S0034425718301792>.
- [48] F. Zellweger, V. Braunisch, F. Morsdorf, A. Baltensweiler, M. Abegg, T. Roth, H. Bugmann e K. Bollmann. “Disentangling the effects of climate, topography, soil and vegetation on stand-scale species richness in temperate forests”. Em: *Forest Ecology and Management* 349 (ago. de 2015), pp. 36–44. DOI: [10.1016/j.foreco.2015.04.008](https://doi.org/10.1016/j.foreco.2015.04.008). URL: <https://www.sciencedirect.com/science/article/pii/S0378112715002029>.
- [49] H. Zhang, M. Huang, X. Qing, G. Li e T. Chuanzhao. “Bibliometric Analysis of Global Remote Sensing Research during 2010–2015”. Em: *ISPRS International Journal of Geo-Information* 6 (nov. de 2017), p. 332. DOI: [10.3390/ijgi6110332](https://doi.org/10.3390/ijgi6110332).



## ANEXO A

Conjunto de Dados	Algoritmo	Classe	Precisão	Recall	F1-score	Kappa
CD_TUDO	KNN	Interven.	0,8333	0,6637	0,7389	0,0000
		Não Interven.	0,9707	0,9882	0,9794	0,0000
		Média	0,9020	0,8260	0,8592	0,7186
	RF	Interven.	0,8713	0,7788	0,8224	0,0000
		Não Interven.	0,9806	0,9898	0,9852	0,0000
		Média	0,9259	0,8843	0,9038	0,8076
	SVM	Interven.	0,7778	0,7434	0,7602	0,0000
		Não Interven.	0,9773	0,9812	0,9793	0,0000
		Média	0,8776	0,8623	0,8697	0,7394
	XGB	Interven.	0,8812	0,7876	0,8318	0,0000
		Não Interven.	0,9814	0,9906	0,9859	0,0000
		Média	0,9313	0,8891	0,9089	0,8178
CD_ÍNDICES	KNN	Interven.	0,7979	0,6637	0,7246	0,0000
		Não Interven.	0,9706	0,9851	0,9778	0,0000
		Média	0,8843	0,8244	0,8512	0,7027
	RF	Interven.	0,8367	0,7257	0,7773	0,0000
		Não Interven.	0,9760	0,9875	0,9817	0,0000
		Média	0,9064	0,8566	0,8795	0,7590
	SVM	Interven.	0,6277	0,7611	0,6880	0,0000
		Não Interven.	0,9784	0,9600	0,9691	0,0000
		Média	0,8031	0,8605	0,8286	0,6574
	XGB	Interven.	0,8265	0,7168	0,7678	0,0000
		Não Interven.	0,9752	0,9867	0,9809	0,0000
		Média	0,9009	0,8517	0,8743	0,7488
CD_BANDAS	KNN	Interven.	0,8333	0,6637	0,7389	0,0000
		Não Interven.	0,9707	0,9882	0,9794	0,0000
		Média	0,9020	0,8260	0,8592	0,7186
	RF	Interven.	0,8252	0,7522	0,7870	0,0000
		Não Interven.	0,9782	0,9859	0,9820	0,0000
		Média	0,9017	0,8690	0,8845	0,7691
	SVM	Interven.	0,7545	0,7345	0,7444	0,0000
		Não Interven.	0,9765	0,9788	0,9777	0,0000
		Média	0,8655	0,8567	0,8610	0,7221
	XGB	Interven.	0,8125	0,6903	0,7464	0,0000
		Não Interven.	0,9729	0,9859	0,9794	0,0000
		Média	0,8927	0,8381	0,8629	0,7259
CD_FUSÃO	KNN	Interven.	0,8571	0,6903	0,7647	0,0000
		Não Interven.	0,9730	0,9898	0,9813	0,0000
		Média	0,9151	0,8400	0,8730	0,7463
	RF	Interven.	0,8454	0,7257	0,7810	0,0000
		Não Interven.	0,9760	0,9882	0,9821	0,0000
		Média	0,9107	0,8569	0,8815	0,7631
	SVM	Interven.	0,6788	0,8230	0,7440	0,0000
		Não Interven.	0,9840	0,9655	0,9747	0,0000
		Média	0,8314	0,8942	0,8593	0,7189
	XGB	Interven.	0,8515	0,7611	0,8037	0,0000
		Não Interven.	0,9790	0,9882	0,9836	0,0000
		Média	0,9153	0,8746	0,8937	0,7874

Figura I.1: Métricas da análise estática para faixas ao longo de estradas.

---

Conjunto de Dados	Algoritmo	FN	FP	TN	TP
CD_TUDO	KNN	38	15	1 260	75
	RF	25	13	1 262	88
	SVM	29	24	1 251	84
	XGB	24	12	1 263	89
CD_ÍNDICES	KNN	38	19	1 256	75
	RF	31	16	1 259	82
	SVM	27	51	1 224	86
	XGB	32	17	1 258	81
CD_BANDAS	KNN	38	15	1 260	75
	RF	28	18	1 257	85
	SVM	30	27	1 248	83
	XGB	35	18	1 257	78
CD_FUSÃO	KNN	35	13	1 262	78
	RF	31	15	1 260	82
	SVM	20	44	1 231	93
	XGB	27	15	1 260	86

Figura I.2: Matriz de confusão da análise estática para faixas ao longo de estradas.

Conjunto de Dados	Algoritmo	Classe	Precisão	Recall	F1-score	Kappa
CD_TUDO	KNN	Interven.	0,7358	0,6393	0,6842	0,0000
		Não Inteven.	0,8896	0,9268	0,9078	0,0000
		Média	0,8127	0,7831	0,7960	0,5926
	RF	Interven.	0,8333	0,6831	0,7508	0,0000
		Não Inteven.	0,9044	0,9564	0,9297	0,0000
		Média	0,8689	0,8198	0,8402	0,6814
	SVM	Interven.	0,6682	0,7705	0,7157	0,0000
		Não Inteven.	0,9231	0,8780	0,9000	0,0000
		Média	0,7957	0,8243	0,8079	0,6164
	XGB	Interven.	0,7935	0,6721	0,7278	0,0000
		Não Inteven.	0,9003	0,9443	0,9218	0,0000
		Média	0,8469	0,8082	0,8248	0,6503
CD_ÍNDICES	KNN	Interven.	0,7278	0,6284	0,6745	0,0000
		Não Inteven.	0,8865	0,9251	0,9054	0,0000
		Média	0,8072	0,7768	0,7899	0,5805
	RF	Interven.	0,7532	0,6503	0,6979	0,0000
		Não Inteven.	0,8932	0,9321	0,9122	0,0000
		Média	0,8232	0,7912	0,8051	0,6107
	SVM	Interven.	0,6580	0,6940	0,6755	0,0000
		Não Inteven.	0,9007	0,8850	0,8928	0,0000
		Média	0,7794	0,7895	0,7842	0,5684
	XGB	Interven.	0,8000	0,6120	0,6935	0,0000
		Não Inteven.	0,8849	0,9512	0,9169	0,0000
		Média	0,8425	0,7816	0,8052	0,6122
CD_BANDAS	KNN	Interven.	0,7375	0,6448	0,6880	0,0000
		Não Inteven.	0,8911	0,9268	0,9086	0,0000
		Média	0,8143	0,7858	0,7983	0,5972
	RF	Interven.	0,7898	0,6776	0,7294	0,0000
		Não Inteven.	0,9017	0,9425	0,9216	0,0000
		Média	0,8457	0,8101	0,8255	0,6516
	SVM	Interven.	0,6698	0,7760	0,7190	0,0000
		Não Inteven.	0,9248	0,8780	0,9008	0,0000
		Média	0,7973	0,8270	0,8099	0,6205
	XGB	Interven.	0,7962	0,6831	0,7353	0,0000
		Não Inteven.	0,9033	0,9443	0,9233	0,0000
		Média	0,8498	0,8137	0,8293	0,6592
CD_FUSÃO	KNN	Interven.	0,7160	0,6339	0,6725	0,0000
		Não Inteven.	0,8874	0,9199	0,9033	0,0000
		Média	0,8017	0,7769	0,7879	0,5763
	RF	Interven.	0,7736	0,6721	0,7193	0,0000
		Não Inteven.	0,8997	0,9373	0,9181	0,0000
		Média	0,8366	0,8047	0,8187	0,6379
	SVM	Interven.	0,6620	0,7814	0,7168	0,0000
		Não Inteven.	0,9261	0,8728	0,8987	0,0000
		Média	0,7940	0,8271	0,8077	0,6164
	XGB	Interven.	0,7736	0,6721	0,7193	0,0000
		Não Inteven.	0,8997	0,9373	0,9181	0,0000
		Média	0,8366	0,8047	0,8187	0,6379

Figura I.3: Métricas da análise estática para faixas ao longo de linhas elétricas.

---

Conjunto de Dados	Algoritmo	FN	FP	TN	TP
CD_TUDO	KNN	66	42	532	117
	RF	58	25	549	125
	SVM	42	70	504	141
	XGB	60	32	542	123
CD_ÍNDICES	KNN	68	43	531	115
	RF	64	39	535	119
	SVM	56	66	508	127
	XGB	71	28	546	112
CD_BANDAS	KNN	65	42	532	118
	RF	59	33	541	124
	SVM	41	70	504	142
	XGB	58	32	542	125
CD_FUSÃO	KNN	67	46	528	116
	RF	60	36	538	123
	SVM	40	73	501	143
	XGB	60	36	538	123

Figura I.4: Matriz de confusão da análise estática para faixas ao longo de linhas elétricas.

Conjunto de Dados	Algoritmo	Classe	Precisão	Recall	F1-score	Kappa
CD_TUDO	KNN	Interven.	0,9263	0,7788	0,8462	0,0000
		Não Interven.	0,9807	0,9945	0,9875	0,0000
		Média	0,9535	0,8866	0,9168	0,8338
	RF	Interven.	0,9223	0,8407	0,8796	0,0000
		Não Interven.	0,9860	0,9937	0,9898	0,0000
		Média	0,9542	0,9172	0,9347	0,8695
	SVM	Interven.	0,8938	0,8938	0,8938	0,0000
		Não Interven.	0,9906	0,9906	0,9906	0,0000
		Média	0,9422	0,9422	0,9422	0,8844
	XGB	Interven.	0,9135	0,8407	0,8756	0,0000
		Não Interven.	0,9860	0,9929	0,9894	0,0000
		Média	0,9497	0,9168	0,9325	0,8650
CD_ÍNDICES	KNN	Interven.	0,7273	0,2832	0,4076	0,0000
		Não Interven.	0,9397	0,9906	0,9645	0,0000
		Média	0,8335	0,6369	0,6861	0,3793
	RF	Interven.	0,9184	0,7965	0,8531	0,0000
		Não Interven.	0,9822	0,9937	0,9879	0,0000
		Média	0,9503	0,8951	0,9205	0,8411
	SVM	Interven.	0,6667	0,5310	0,5911	0,0000
		Não Interven.	0,9592	0,9765	0,9677	0,0000
		Média	0,8129	0,7537	0,7794	0,5593
	XGB	Interven.	0,8857	0,8230	0,8532	0,0000
		Não Interven.	0,9844	0,9906	0,9875	0,0000
		Média	0,9351	0,9068	0,9204	0,8407
CD_BANDAS	KNN	Interven.	0,9255	0,7699	0,8406	0,0000
		Não Interven.	0,9799	0,9945	0,9872	0,0000
		Média	0,9527	0,8822	0,9139	0,8279
	RF	Interven.	0,9300	0,8230	0,8732	0,0000
		Não Interven.	0,9845	0,9945	0,9895	0,0000
		Média	0,9572	0,9088	0,9314	0,8627
	SVM	Interven.	0,8793	0,9027	0,8908	0,0000
		Não Interven.	0,9914	0,9890	0,9902	0,0000
		Média	0,9353	0,9458	0,9405	0,8810
	XGB	Interven.	0,8972	0,8496	0,8727	0,0000
		Não Interven.	0,9867	0,9914	0,9890	0,0000
		Média	0,9420	0,9205	0,9309	0,8618
CD_FUSÃO	KNN	Interven.	0,9158	0,7699	0,8365	0,0000
		Não Interven.	0,9799	0,9937	0,9868	0,0000
		Média	0,9478	0,8818	0,9116	0,8234
	RF	Interven.	0,9320	0,8496	0,8889	0,0000
		Não Interven.	0,9868	0,9945	0,9906	0,0000
		Média	0,9594	0,9220	0,9398	0,8795
	SVM	Interven.	0,8661	0,8584	0,8622	0,0000
		Não Interven.	0,9875	0,9882	0,9878	0,0000
		Média	0,9268	0,9233	0,9250	0,8501
	XGB	Interven.	0,9314	0,8407	0,8837	0,0000
		Não Interven.	0,9860	0,9945	0,9902	0,0000
		Média	0,9587	0,9176	0,9370	0,8740

Figura I.5: Métricas da análise temporal para faixas ao longo de estradas usando dados do Sentinel-2.



Conjunto de Dados	Algoritmo	FN	FP	TN	TP
CD_TUDO	KNN	25	7	1 268	88
	RF	18	8	1 267	95
	SVM	12	12	1 263	101
	XGB	18	9	1 266	95
CD_ÍNDICES	KNN	81	12	1 263	32
	RF	23	8	1 267	90
	SVM	53	30	1 245	60
	XGB	20	12	1 263	93
CD_BANDAS	KNN	26	7	1 268	87
	RF	20	7	1 268	93
	SVM	11	14	1 261	102
	XGB	17	11	1 264	96
CD_FUSÃO	KNN	26	8	1 267	87
	RF	17	7	1 268	96
	SVM	16	15	1 260	97
	XGB	18	7	1 268	95

Figura I.6: Matriz de confusão da análise temporal para faixas ao longo de estradas usando dados do Sentinel-2.

Conjunto de Dados (VV, VH, VV-VH, VV/VH)	Algoritmo	Classe	Precisão	Recall	F1-score	Kappa
	KNN	Interven.	0,571	0,035	0,067	0,000
		Não Interven.	0,921	0,998	0,958	0,000
		Média	0,746	0,517	0,512	0,058
	RF	Interven.	0,559	0,292	0,384	0,000
		Não Interven.	0,940	0,980	0,959	0,000
		Média	0,750	0,636	0,672	0,347
	SVM	Interven.	0,000	0,000	0,000	0,000
		Não Interven.	0,919	1,000	0,958	0,000
		Média	0,459	0,500	0,479	0,000
	XGB	Interven.	0,535	0,204	0,295	0,000
		Não Interven.	0,933	0,984	0,958	0,000
		Média	0,734	0,594	0,626	0,262

Figura I.7: Métricas da análise temporal para faixas ao longo de estradas usando dados do Sentinel-1.

Conjunto de Dados	Algoritmo	FN	FP	TN	TP
(VV, VH, VV-VH, VV/VH)	KNN	109	3	1 272	4
	RF	80	26	1 249	33
	SVM	113		1 275	
	XGB	90	20	1 255	23

Figura I.8: Matriz de confusão da análise temporal para faixas ao longo de estradas usando dados do Sentinel-1.

Conjunto de Dados	Algoritmo	Classe	Precisão	Recall	F1-score	Kappa
CD_TUDO	KNN	Interven.	0,9271	0,7876	0,8517	0,0000
		Não Interven.	0,9814	0,9945	0,9879	0,0000
		Média	0,9543	0,8911	0,9198	0,8397
	RF	Interven.	0,9307	0,8319	0,8785	0,0000
		Não Interven.	0,9852	0,9945	0,9899	0,0000
		Média	0,9580	0,9132	0,9342	0,8684
	SVM	Interven.	0,8938	0,8938	0,8938	0,0000
		Não Interven.	0,9906	0,9906	0,9906	0,0000
		Média	0,9422	0,9422	0,9422	0,8844
	XGB	Interven.	0,9320	0,8496	0,8889	0,0000
		Não Interven.	0,9868	0,9945	0,9906	0,0000
		Média	0,9594	0,9220	0,9398	0,8795
CD_ÍNDICES	KNN	Interven.	0,7069	0,3628	0,4795	0,0000
		Não Interven.	0,9459	0,9867	0,9658	0,0000
		Média	0,8264	0,6747	0,7227	0,4491
	RF	Interven.	0,9053	0,7611	0,8269	0,0000
		Não Interven.	0,9791	0,9929	0,9860	0,0000
		Média	0,9422	0,8770	0,9065	0,8130
	SVM	Interven.	0,6588	0,4956	0,5657	0,0000
		Não Interven.	0,9563	0,9773	0,9666	0,0000
		Média	0,8075	0,7364	0,7661	0,5330
	XGB	Interven.	0,9192	0,8053	0,8585	0,0000
		Não Interven.	0,9829	0,9937	0,9883	0,0000
		Média	0,9511	0,8995	0,9234	0,8468
CD_BANDAS	KNN	Interven.	0,9263	0,7788	0,8462	0,0000
		Não Interven.	0,9807	0,9945	0,9875	0,0000
		Média	0,9535	0,8866	0,9168	0,8338
	RF	Interven.	0,9126	0,8319	0,8704	0,0000
		Não Interven.	0,9852	0,9929	0,9891	0,0000
		Média	0,9489	0,9124	0,9297	0,8595
	SVM	Interven.	0,8793	0,9027	0,8908	0,0000
		Não Interven.	0,9914	0,9890	0,9902	0,0000
		Média	0,9353	0,9458	0,9405	0,8810
	XGB	Interven.	0,8846	0,8142	0,8479	0,0000
		Não Interven.	0,9836	0,9906	0,9871	0,0000
		Média	0,9341	0,9024	0,9175	0,8351
CD_FUSÃO	KNN	Interven.	0,9255	0,7699	0,8406	0,0000
		Não Interven.	0,9799	0,9945	0,9872	0,0000
		Média	0,9527	0,8822	0,9139	0,8279
	RF	Interven.	0,9490	0,8230	0,8815	0,0000
		Não Interven.	0,9845	0,9961	0,9903	0,0000
		Média	0,9667	0,9095	0,9359	0,8718
	SVM	Interven.	0,8661	0,8584	0,8622	0,0000
		Não Interven.	0,9875	0,9882	0,9878	0,0000
		Média	0,9268	0,9233	0,9250	0,8501
	XGB	Interven.	0,9200	0,8142	0,8638	0,0000
		Não Interven.	0,9837	0,9937	0,9887	0,0000
		Média	0,9518	0,9039	0,9263	0,8526

Figura I.9: Métricas da análise temporal para faixas ao longo de estradas usando dados do Sentinel-2 e Sentinel-1.

Conjunto de Dados	Algoritmo	FN	FP	TN	TP
CD_TUDO	KNN	24	7	1 268	89
	RF	19	7	1 268	94
	SVM	12	12	1 263	101
	XGB	17	7	1 268	96
CD_ÍNDICES	KNN	72	17	1 258	41
	RF	27	9	1 266	86
	SVM	57	29	1 246	56
	XGB	22	8	1 267	91
CD_BANDAS	KNN	25	7	1 268	88
	RF	19	9	1 266	94
	SVM	11	14	1 261	102
	XGB	21	12	1 263	92
CD_FUSÃO	KNN	26	7	1 268	87
	RF	20	5	1 270	93
	SVM	16	15	1 260	97
	XGB	21	8	1 267	92

Figura I.10: Matriz de confusão da análise temporal para faixas ao longo de estradas usando dados do Sentinel-2 e Sentinel-1.

Conjunto de Dados	Algoritmo	Classe	Precisão	Recall	F1-score	Kappa
CD_TUDO	KNN	Interven.	0,8026	0,6667	0,7284	0,0000
		Não Interven.	0,8992	0,9477	0,9228	0,0000
		Média	0,8509	0,8072	0,8256	0,6520
	RF	Interven.	0,8599	0,7377	0,7941	0,0000
		Não Interven.	0,9200	0,9617	0,9404	0,0000
		Média	0,8899	0,8497	0,8672	0,7349
	SVM	Interven.	0,7566	0,7814	0,7688	0,0000
		Não Interven.	0,9296	0,9199	0,9247	0,0000
		Média	0,8431	0,8506	0,8468	0,6935
	XGB	Interven.	0,8313	0,7541	0,7908	0,0000
		Não Interven.	0,9239	0,9512	0,9373	0,0000
		Média	0,8776	0,8527	0,8641	0,7284
CD_ÍNDICES	KNN	Interven.	0,7778	0,5738	0,6604	0,0000
		Não Interven.	0,8746	0,9477	0,9097	0,0000
		Média	0,8262	0,7608	0,7850	0,5727
	RF	Interven.	0,8225	0,7596	0,7898	0,0000
		Não Interven.	0,9252	0,9477	0,9363	0,0000
		Média	0,8738	0,8536	0,8630	0,7262
	SVM	Interven.	0,7193	0,4481	0,5522	0,0000
		Não Interven.	0,8429	0,9443	0,8907	0,0000
		Média	0,7811	0,6962	0,7215	0,4501
	XGB	Interven.	0,8457	0,7486	0,7942	0,0000
		Não Interven.	0,9227	0,9564	0,9393	0,0000
		Média	0,8842	0,8525	0,8667	0,7338
CD_BANDAS	KNN	Interven.	0,8026	0,6667	0,7284	0,0000
		Não Interven.	0,8992	0,9477	0,9228	0,0000
		Média	0,8509	0,8072	0,8256	0,6520
	RF	Interven.	0,8125	0,7104	0,7580	0,0000
		Não Interven.	0,9112	0,9477	0,9291	0,0000
		Média	0,8619	0,8291	0,8436	0,6875
	SVM	Interven.	0,7500	0,7869	0,7680	0,0000
		Não Interven.	0,9310	0,9164	0,9236	0,0000
		Média	0,8405	0,8516	0,8458	0,6917
	XGB	Interven.	0,9085	0,7049	0,7938	0,0000
		Não Interven.	0,9122	0,9774	0,9437	0,0000
		Média	0,9103	0,8411	0,8687	0,7386
CD_FUSÃO	KNN	Interven.	0,8105	0,6776	0,7381	0,0000
		Não Interven.	0,9023	0,9495	0,9253	0,0000
		Média	0,8564	0,8135	0,8317	0,6642
	RF	Interven.	0,8516	0,7213	0,7811	0,0000
		Não Interven.	0,9153	0,9599	0,9371	0,0000
		Média	0,8834	0,8406	0,8591	0,7187
	SVM	Interven.	0,7296	0,7814	0,7546	0,0000
		Não Interven.	0,9287	0,9077	0,9181	0,0000
		Média	0,8291	0,8445	0,8363	0,6728
	XGB	Interven.	0,8533	0,6995	0,7688	0,0000
		Não Interven.	0,9094	0,9617	0,9348	0,0000
		Média	0,8814	0,8306	0,8518	0,7044

Figura I.11: Métricas da análise temporal para faixas ao longo de linhas usando dados do Sentinel-2.

Conjunto de Dados	Algoritmo	FN	FP	TN	TP
CD_TUDO	KNN	61	30	544	122
	RF	48	22	552	135
	SVM	40	46	528	143
	XGB	45	28	546	138
CD_ÍNDICES	KNN	78	30	544	105
	RF	44	30	544	139
	SVM	101	32	542	82
	XGB	46	25	549	137
CD_BANDAS	KNN	61	30	544	122
	RF	53	30	544	130
	SVM	39	48	526	144
	XGB	54	13	561	129
CD_FUSÃO	KNN	59	29	545	124
	RF	51	23	551	132
	SVM	40	53	521	143
	XGB	55	22	552	128

Figura I.12: Matriz de confusão da análise temporal para faixas ao longo de linhas usando dados do Sentinel-2.

Conjunto de Dados	Algoritmo	Classe	Precisão	Recall	F1-score	Kappa
(VV, VH, VV-VH, VV/VH)	KNN	Interven.	0,6569	0,3661	0,4702	0,0000
		Não Interven.	0,8229	0,9390	0,8771	0,0000
		Média	0,7399	0,6526	0,6737	0,3593
	RF	Interven.	0,7170	0,6230	0,6667	0,0000
		Não Interven.	0,8846	0,9216	0,9027	0,0000
		Média	0,8008	0,7723	0,7847	0,5700
	SVM	Interven.	0,6629	0,3224	0,4338	0,0000
		Não Interven.	0,8144	0,9477	0,8760	0,0000
		Média	0,7386	0,6351	0,6549	0,3274
	XGB	Interven.	0,7402	0,5137	0,6065	0,0000
		Não Interven.	0,8587	0,9425	0,8987	0,0000
		Média	0,7994	0,7281	0,7526	0,5092

Figura I.13: Métricas da análise temporal para faixas ao longo de linhas usando dados do Sentinel-1.

---

Conjunto de Dados	Algoritmo	FN	FP	TN	TP
(VV, VH, VV-VH, VV/VH)	KNN	116	35	539	67
	RF	69	45	529	114
	SVM	124	30	544	59
	XGB	89	33	541	94

Figura I.14: Matriz de confusão da análise temporal para faixas ao longo de linhas usando dados do Sentinel-1.

Conjunto de Dados	Algoritmo	Classe	Precisão	Recall	F1-score	Kappa
CD_TUDO	KNN	Interven.	0,8026	0,6667	0,7284	0,0000
		Não Interven.	0,8992	0,9477	0,9228	0,0000
		Média	0,8509	0,8072	0,8256	0,6520
	RF	Interven.	0,8581	0,7268	0,7870	0,0000
		Não Interven.	0,9169	0,9617	0,9388	0,0000
		Média	0,8875	0,8442	0,8629	0,7263
	SVM	Interven.	0,7566	0,7814	0,7688	0,0000
		Não Interven.	0,9296	0,9199	0,9247	0,0000
		Média	0,8431	0,8506	0,8468	0,6935
	XGB	Interven.	0,8500	0,7432	0,7930	0,0000
		Não Interven.	0,9213	0,9582	0,9394	0,0000
		Média	0,8856	0,8507	0,8662	0,7327
CD_ÍNDICES	KNN	Interven.	0,7810	0,5847	0,6687	0,0000
		Não Interven.	0,8774	0,9477	0,9112	0,0000
		Média	0,8292	0,7662	0,7900	0,5823
	RF	Interven.	0,8405	0,7486	0,7919	0,0000
		Não Interven.	0,9226	0,9547	0,9384	0,0000
		Média	0,8815	0,8517	0,8651	0,7305
	SVM	Interven.	0,7211	0,5792	0,6424	0,0000
		Não Interven.	0,8738	0,9286	0,9003	0,0000
		Média	0,7974	0,7539	0,7714	0,5443
	XGB	Interven.	0,8133	0,7377	0,7736	0,0000
		Não Interven.	0,9188	0,9460	0,9322	0,0000
		Média	0,8660	0,8418	0,8529	0,7060
CD_BANDAS	KNN	Interven.	0,8026	0,6667	0,7284	0,0000
		Não Interven.	0,8992	0,9477	0,9228	0,0000
		Média	0,8509	0,8072	0,8256	0,6520
	RF	Interven.	0,8408	0,7213	0,7765	0,0000
		Não Interven.	0,9150	0,9564	0,9353	0,0000
		Média	0,8779	0,8389	0,8559	0,7122
	SVM	Interven.	0,7539	0,7869	0,7701	0,0000
		Não Interven.	0,9311	0,9181	0,9246	0,0000
		Média	0,8425	0,8525	0,8473	0,6947
	XGB	Interven.	0,8497	0,7104	0,7738	0,0000
		Não Interven.	0,9123	0,9599	0,9355	0,0000
		Média	0,8810	0,8352	0,8546	0,7100
CD_FUSÃO	KNN	Interven.	0,8105	0,6776	0,7381	0,0000
		Não Interven.	0,9023	0,9495	0,9253	0,0000
		Média	0,8564	0,8135	0,8317	0,6642
	RF	Interven.	0,8452	0,7158	0,7751	0,0000
		Não Interven.	0,9136	0,9582	0,9354	0,0000
		Média	0,8794	0,8370	0,8553	0,7111
	SVM	Interven.	0,7268	0,7268	0,7268	0,0000
		Não Interven.	0,9129	0,9129	0,9129	0,0000
		Média	0,8198	0,8198	0,8198	0,6397
	XGB	Interven.	0,8188	0,7158	0,7638	0,0000
		Não Interven.	0,9129	0,9495	0,9308	0,0000
		Média	0,8658	0,8327	0,8473	0,6951

Figura I.15: Métricas da análise temporal para faixas ao longo de linhas usando dados do Sentinel-2 e Sentinel-1.



---

Conjunto de Dados	Algoritmo	FN	FP	TN	TP
CD_TUDO	KNN	61	30	544	122
	RF	50	22	552	133
	SVM	40	46	528	143
	XGB	47	24	550	136
CD_ÍNDICES	KNN	76	30	544	107
	RF	46	26	548	137
	SVM	77	41	533	106
	XGB	48	31	543	135
CD_BANDAS	KNN	61	30	544	122
	RF	51	25	549	132
	SVM	39	47	527	144
	XGB	53	23	551	130
CD_FUSÃO	KNN	59	29	545	124
	RF	52	24	550	131
	SVM	50	50	524	133
	XGB	52	29	545	131

Figura I.16: Matriz de confusão da análise temporal para faixas ao longo de linhas usando dados do Sentinel-2 e e Sentinel-1.